



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 169185

TO: Jeanine Goldberg
Location: REM-2D15&2C70
Art Unit: 1634
Monday, October 31, 2005
Case Serial Number: 10/654416

From: Toby Port
Location: Biotech-Chem Library
REM-1A59
Phone: 571-272-2523

toby.port@uspto.gov

Search Notes

Examiner Goldberg,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Toby Port
X22523

This Page Blank (uspto)



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or *contact*:

Mary Hale, Information Branch Supervisor
Remsen Bldg. 01 D86
571-272-2507

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC-Biotech-Chem Library, Remsen Bldg.



This Page Blank (uspto)

STIC-Biotech/ChemLib

169/86

From: Goldberg, Jeanine
Sent: Thursday, October 20, 2005 2:42 PM
To: STIC-Biotech/ChemLib
Subject: 10/654,416- Search

Hello- Please search for a nucleic acid encoding SEQ ID NO: 14.

Please search for SEQ ID NO: 13.

THANK YOU

Jeanine Anne Goldberg
1634
571-272-0743
REM 2D15
Mailbox: 2C70

CRFS

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

This Page Blank (uspio)

241 ATTGAAAAATGTATCAATCAAGAAAACTTAAAGAACTTAAAAAAATTT 300
301 CAGATTTAACGATPAAGATTGCAAGAGAAATGAACAACAATGCTATTTTGGAGGA 360
301 CAGATTTAACGATPAAGATTGCAAGAGAAATGAACAACAATGCTATTTTGGAGGA 360
361 GTATGTTCAAAAGAACTTAAAGATGATGCAATTAAGATTAAGTCTATCAAAAG 420
361 GTATGTTCAAAAGAACTTAAAGATGATGCAATTAAGATTAAGTCTATCAAAAG 420
421 AAACGTTAAGTTGCGGAGAAAGTTCTTTAAGAGACTTCGAGAGATCTTAAATGA 480
421 AAACGTTAAGTTGCGGAGAAAGTTCTTTAAGAGACTTCGAGAGATCTTAAATGA 480
481 TCAGTCATATGTGAATAAACTTAAAGAGATTTCCCTGTCATGGGAGGAAAGTAT 540
481 TCAGTCATATGTGAATAAACTTAAAGAGATTTCCCTGTCATGGGAGGAAAGTAT 540
541 GAGTTAACAACTTGTCGTGAACAGAAAGAGACATGTAAAGATATTTAATTGAAAA 600
541 GAGTTAACAACTTGTCGTGAACAGAAAGAGACATGTAAAGATATTTAATTGAAAA 600
601 GATAAAGAGTGGGATCTTAAACAGATGTTTCAAGACACTAGAAAGTTTAAAAA 660
601 GATAAAGAGTGGGATCTTAAACAGATGTTTCAAGACACTAGAAAGTTTAAAAA 660
661 GAAACATGCTTGAATTAATCGAACAATGCTATTTTCAATGAAATTCGAGACGAC 720
661 GAAACATGCTTGAATTAATCGAACAATGCTATTTTCAATGAAATTCGAGACGAC 720
721 GATATTAATTAATGTATTAATGAGAGAGAAATCCCAAGAACAAATATGCTATATG 780
721 GATATTAATTAATGTATTAATGAGAGAGAAATCCCAAGAACAAATATGCTATATG 780
781 CCACGAGACCCGATTTTGAATCCACTAGCCAGAGGCTACATAGCAGAGATTAAGG 840
781 CCACGAGACCCGATTTTGAATCCACTAGCCAGAGGCTACATAGCAGAGATTAAGG 840
841 CTGGAAGAGTTTATPAAGAGATGAGAGAGATGAGTTTATTTGAAAGATCATCTA 900
841 CTGGAAGAGTTTATPAAGAGATGAGAGAGATGAGTTTATTTGAAAGATCATCTA 900
901 AGAGATGCGACAGCTTTGTTGGCATTTGATCCAAAGATTCATGCTTAAAAAAGAC 960
901 AGAGATGCGACAGCTTTGTTGGCATTTGATCCAAAGATTCATGCTTAAAAAAGAC 960
961 GACAAAGAGAAATGCGAAGAGCCCTTCAAAAACCTGCAAAAATCTCATGAACATGAG 1020
961 GACAAAGAGAAATGCGAAGAGCCCTTCAAAAACCTGCAAAAATCTCATGAACATGAG 1020
1021 GCTTTAGAAAGTTTATPAAGAGATGAGAGAGATGAGTTTATTTGAAAGATCATCTA 1080
1021 GCTTTAGAAAGTTTATPAAGAGATGAGAGAGATGAGTTTATTTGAAAGATCATCTA 1080
1081 GAATTTGCAAAATGATATTAACAACTTGCAAAATTTTCACTTCAAAAGCTCATATAT 1140
1081 GAATTTGCAAAATGATATTAACAACTTGCAAAATTTTCACTTCAAAAGCTCATATAT 1140
1141 CGTCTTTTGAATTCACAAAGAAATTAAGAAATTTGTTGATGAGAGAGGTTGCAACA 1200
1141 CGTCTTTTGAATTCACAAAGAAATTAAGAAATTTGTTGATGAGAGAGGTTGCAACA 1200
1201 TTTCTTGAAGAGATTTGTCGAAATTTGAGTCTTATTTGTTTCAATTTGAAAAA 1260
1201 TTTCTTGAAGAGATTTGTCGAAATTTGAGTCTTATTTGTTTCAATTTGAAAAA 1260
1261 TGTCCAGATGAGAGAAATGATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
1261 TGTCCAGATGAGAGAAATGATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
1321 GATGACGGGCAATTAAGTGTCTGCAAGAAATATGCGAGAGATGTACATGTTCAAC 1380
1321 GATGACGGGCAATTAAGTGTCTGCAAGAAATATGCGAGAGATGTACATGTTCAAC 1380

1321 GATGACGGGCAATTAAGTGTCTGCAAGAAATATGCGAGAGATGTACATGTTCAAC 1380
1381 AAAAGTGGCTTGAAGAGTTTCAACAAGATTTAGTAAAGATATGAGAGAGAGAGAGAG 1440
1381 AAAAGTGGCTTGAAGAGTTTCAACAAGATTTAGTAAAGATATGAGAGAGAGAGAGAG 1440
1441 GAAACCAAGAGAGTTTCTCAAGAGATGATATTTATTTATTTGTTGTTGTTGTTGTTGTT 1500
1441 GAAACCAAGAGAGTTTCTCAAGAGATGATATTTATTTATTTGTTGTTGTTGTTGTTGTT 1500
1501 GCAGCCGGTTGCTTACAGATGATCTTGAATGAAATCTATCTTTTACAGCAACATCTG 1560
1501 GCAGCCGGTTGCTTACAGATGATCTTGAATGAAATCTATCTTTTACAGCAACATCTG 1560
1561 GATCAAAAGCAGATTTCCGACAGATTAATAATGCAAGAAATTTGGGAGAGAGAGAGAG 1620
1561 GATCAAAAGCAGATTTCCGACAGATTAATAATGCAAGAAATTTGGGAGAGAGAGAGAG 1620
1621 GATTTAGAGAGAGTTTCAAGAGATTAATGAGAGATGATGATGATGATGATGATGATGAT 1680
1621 GATTTAGAGAGAGTTTCAAGAGATTAATGAGAGATGATGATGATGATGATGATGATGAT 1680
1681 AATGCTTGGGAGCTACAGAAATTTTAAAGCAGGTTTATTTGATGAGACAGAAAGATCT 1740
1681 AATGCTTGGGAGCTACAGAAATTTTAAAGCAGGTTTATTTGATGAGACAGAAAGATCT 1740
1741 TTGAAGACCAAGAAAGTTGTGTAAAAATCCTAAAGAAAGATGATATTAATGATGATGAT 1800
1741 TTGAAGACCAAGAAAGTTGTGTAAAAATCCTAAAGAAAGATGATATTAATGATGATGAT 1800
1801 AGAGAGATGACGTTTCTCTTTTGTATGTTCTTTCCAAACGCTACGTTGAGAGTGTG 1860
1801 AGAGAGATGACGTTTCTCTTTTGTATGTTCTTTCCAAACGCTACGTTGAGAGTGTG 1860
1801 AGAGAGATGACGTTTCTCTTTTGTATGTTCTTTCCAAACGCTACGTTGAGAGTGTG 1860
1801 AGAGAGATGACGTTTCTCTTTTGTATGTTCTTTCCAAACGCTACGTTGAGAGTGTG 1860
1861 GTAAAGACGTAAGAGACAGGTTGTAAGTATTCAAAAAATATTAAGTTCAATATTT 1920
1861 GTAAAGACGTAAGAGACAGGTTGTAAGTATTCAAAAAATATTAAGTTCAATATTT 1920
1861 GTAAAGACGTAAGAGACAGGTTGTAAGTATTCAAAAAATATTAAGTTCAATATTT 1920
1861 GTAAAGACGTAAGAGACAGGTTGTAAGTATTCAAAAAATATTAAGTTCAATATTT 1920
1921 ATTGAATTTCTTGAAGAAATTAATCAAAATTAATAAACAACCTGAAAGAAATTTGCCCT 1980
1921 ATTGAATTTCTTGAAGAAATTAATCAAAATTAATAAACAACCTGAAAGAAATTTGCCCT 1980
1981 TGGCATACGATTTTGCATTAATTTTCACTTAATTTGTCAGAGTCTTACGAAAGAGATGAT 2040
1981 TGGCATACGATTTTGCATTAATTTTCACTTAATTTGTCAGAGTCTTACGAAAGAGATGAT 2040
1981 TGGCATACGATTTTGCATTAATTTTCACTTAATTTGTCAGAGTCTTACGAAAGAGATGAT 2040
2041 TGTCAAAATCAAGAGATGAGAGATGAGAGAGGCTTCTAATAAAGAGAGGCTTGAAGATCT 2100
2041 TGTCAAAATCAAGAGATGAGAGATGAGAGAGGCTTCTAATAAAGAGAGGCTTGAAGATCT 2100
2101 CTCAAGTGAAGCTTCAAGAGAAATTTGAATGATTAATCTTAATGATGAACTTGCAATTGAAA 2160
2101 CTCAAGTGAAGCTTCAAGAGAAATTTGAATGATTAATCTTAATGATGAACTTGCAATTGAAA 2160
2161 AGATATTTGTACAGTACGAGAGAGAGAGATTAATGATGATGATGATGATGATGATGATGAT 2220
2161 AGATATTTGTACAGTACGAGAGAGAGAGATTAATGATGATGATGATGATGATGATGATGAT 2220
2221 AAACCAAGAGATTAATCTGGAAGAGTGAAGAGATGATGATGATGATGATGATGATGATGAT 2280
2221 AAACCAAGAGATTAATCTGGAAGAGTGAAGAGATGATGATGATGATGATGATGATGATGAT 2280
2281 TTAAGTGAAGAGAGTGAAGAGAGTGAAGAGATTAATGATGATGATGATGATGATGATGATGAT 2340
2281 TTAAGTGAAGAGAGTGAAGAGAGTGAAGAGATTAATGATGATGATGATGATGATGATGATGAT 2340
2341 GCTGATCTTAAAAAAGATTTATTAAGATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
2341 GCTGATCTTAAAAAAGATTTATTAAGATATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
2401 ATGAACAGTCCAGCTTTGTTGTCATCTCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
2401 ATGAACAGTCCAGCTTTGTTGTCATCTCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460

QY 2461 AGTAATAGCAAAAACAGATPAAGATGCGTTTCAAAACGAGCTTCAAGATACCAAAA 2520
 DB 2461 AGTAATAGCAAAAACAGATPAAGATGCGTTTCAAAACGAGCTTCAAGATACCAAAA 2520
 QY 2521 CATGTGAAAATCTACGAGAGGGAGTTAAGATGTATCCGTACAGAAATTGAAGCTAAA 2580
 DB 2521 CATGTGAAAATCTACGAGAGGGAGTTAAGATGTATCCGTACAGAAATTGAAGCTAAA 2580
 QY 2581 GCATTGTATTGGCAGCAGAAAGTATTGGAAAGATGTGAAAGTAAAGTAAAT 2640
 DB 2581 GCATTGTATTGGCAGCAGAAAGTATTGGAAAGATGTGAAAGTAAAGTAAAT 2640
 QY 2641 AAATTGGATCAGATTGAGAAATTAGAGAGATTGCAAAAGACTTAAAGAAATGTGAAA 2700
 DB 2641 AAATTGGATCAGATTGAGAAATTAGAGAGATTGCAAAAGACTTAAAGAAATGTGAAA 2700
 QY 2701 AAGATTATATAGGCTTGTCTGCAATCTGAAGCTCTGAGAGTAAAGCCGACGAAACAGTG 2760
 DB 2701 AAGATTATATAGGCTTGTCTGCAATCTGAAGCTCTGAGAGTAAAGCCGACGAAACAGTG 2760
 QY 2761 ACAGAAAAGTACAGCACTACCAACAAACAAACACCCGTTGCCGATCCGAAAGCAACG 2820
 DB 2761 ACAGAAAAGTACAGCACTACCAACAAACAAACACCCGTTGCCGATCCGAAAGCAACG 2820
 QY 2821 GAATGCAAAATCCTTACAGCAACAGACATGAGGTTTACAGAGATGAGACACACAAAGC 2880
 DB 2821 GAATGCAAAATCCTTACAGCAACAGACATGAGGTTTACAGAGATGAGACACACAAAGC 2880
 QY 2881 ACCTCTACTATCAGATCTACCATCATCAATCAAAAATTAACATTGACATCAACGAGCGATGC 2940
 DB 2881 ACCTCTACTATCAGATCTACCATCATCAATCAAAAATTAACATTGACATCAACGAGCGATGC 2940
 QY 2941 AAACCAACCAAGTGTACGACGAGAGAGAGATGATGACGAGACGCTGAAAACCGAGTGAG 3000
 DB 2941 AAACCAACCAAGTGTACGACGAGAGAGAGATGATGACGAGACGCTGAAAACCGAGTGAG 3000
 QY 3001 GGGGTGAGATGAGAGGGTGAATGTGATGAGGGGGGATGATGATGAGAAATGTTTTCG 3060
 DB 3001 GGGGTGAGATGAGAGGGTGAATGTGATGAGGGGGGATGATGATGAGAAATGTTTTCG 3060
 QY 3061 TTTCATGATTTAG 3072
 DB 3061 TTTCATGATTTAG 3072

RESULT 2
 AF033211 3071 bp DNA linear ~~PLN 10-SEP-1998~~
 LOCUS Pneumocystis carinii f. sp. hominis clone HumSG35 major surface
 DEFINITION glycoprotein (MSG) gene, partial cde.
 ACCESSION AF033211 GI:3560518
 VERSION AF033211.1
 KEYWORDS
 SOURCE
 ORGANISM
 Pneumocystis jirovecii
 Pneumocystis jirovecii
 Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
 Pneumocystidaceae; Pneumocystis.
 REFERENCE
 AUTHORS Garbe, T.R. and Striinger, J.R.
 TITLE Molecular characterization of clustered variants of genes encoding
 major surface antigens of human Pneumocystis carinii
 JOURNAL Infect. Immun. 62 (6), 3092-3101 (1994)
 MEDLINE 94314421
 PUBMED 751806
 REFERENCE
 AUTHORS Mei, Q., Turner, R.E., Social, V., Klivington, D., Angus, C.W. and
 TITLE Characterization of major surface glycoprotein genes of human
 Pneumocystis carinii and high-level expression of a conserved
 region
 JOURNAL Infect. Immun. 66 (9), 4268-4273 (1998)
 MEDLINE 98380374

PUBMED 9712777
 REFERENCE 3 (bases 1 to 3071)
 AUTHORS Mei, Q., Turner, R., Social, V., Klivington, D., Angus, C.W. and
 TITLE Direct Submission
 JOURNAL Submitted (07-NOV-1997) Critical Care Medicine Dept., National
 Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD
 20892-1662, USA

FEATURES
 source
 location/Qualifiers
 1..3071
 /organism="Pneumocystis jirovecii"
 /mol_type="genomic DNA"
 /specific_host="Homo sapiens"
 /sub_species="hominis"
 /db_xref="taxon:42068"
 /clone="HumSG35"
 /note="derived from HIV-infected human with P. carinii
 pneumonia"
 <1..3071
 /gene="MSG"
 /note="surface antigen"
 /codon_start=3
 /product="major surface glycoprotein"
 /protein_id="AAC34974.1"
 /db_xref="GI:3560519"
 /translation="AAAVKQAAAGTQNSIDBEHYALILKEDGSLBDECKKLLKXCO
 ELTEAKUNITQVHRKLPCEDEGADPTCKELKANIIEKCTTIGKLEIKKIQTIT
 TDQCKENBOQCLFLEGVCSKELKDDNTLNKCYOKRQKVAEVLRLRLSDNS
 VICEKLEKEICPVMGRESDELITNLCLNOKETCKNIIEXPKCGTLLTDVSAALGFR
 KENCLLEBOCYFYI GNCGDDIYKCELGKCOBOJIAVMPDPDPYRPAETIAE
 DIGLEFVKVEEDGVIFGNHLDALFALLIIDSLSLKDKDKCEBCEALKNENY
 PHEBEALESI CKKNGLSNDGTCKCEBLNDINTCKI FTSKVTNNRLPDPKKNENY
 GWEGLPTFLSNEQCAKLESYCFYPERKCPGENACNIRATCTYKRGIDAPANKVLOEN
 MRGLHNSNSWLEKFOQLVYKCEKCKENKKSFSNDELFLCVQAKAARLTHLD
 RKMTIFLRQOLDQKRPDPIDNKCELEKRCODIGESKEITWPCHTLEOCCNRIGTTE
 ILKOVLLDEHKDITLKDESCVKYKCEKCNKSRGDDRFSPVCFQNAETCELANVDYK
 EAKAFDLAAEVFGRYVDLKERCNKLESDCKIKEDCKDLCEVCKINKACNKLKLEVK
 PHEVTETSTTTTTTTTVAADPKATECKSLQTDYVTWOTSTSTSTSTSTSTSTST
 LTSTRCKPKTKCTTGGEDDAGDVKPSGLMSGNWVRGVI VAAVVISFMT"

ORIGIN
 Query Match 100.0%; Score 3071; DB 8; Length 3071;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 TGCGCGCGCGCGGTAAACGGCAGCAGGACACAGATGATGATGAGAAACGTG 61
 DB 1 TGCGCGCGCGCGGTAAACGGCAGCAGGACACAGATGATGATGAGAAACGTG 60
 QY 62 TTTTAAAGTTTAAATCTTAAAGGAAGATGAGTAAATGAGCAGAAATGCAAAAAAAGCTAA 121
 DB 61 TTTTAAAGTTTAAATCTTAAAGGAAGATGAGTAAATGAGCAGAAATGCAAAAAAAGCTAA 120
 QY 122 AAAAATATTTGCAAGAAATTGACTGAACAAAACCTAAATATAGAACAAATGACAGAAAAC 181
 DB 121 AAAAATATTTGCAAGAAATTGACTGAACAAAACCTAAATATAGAACAAATGACAGAAAAC 180
 QY 182 TTTAAAGTTTTTTGGGAAAGATGAAAAACAGATACAAATAATGCAAGAACTGAAGACCATATA 241
 DB 181 TTTAAAGTTTTTTGGGAAAGATGAAAAACAGATACAAATAATGCAAGAACTGAAGACCATATA 240
 QY 242 TTGAGAAAAATGTAATCTACATCAAGGAAAACTTAAAGAAAGCAATTTAAAAAAATTTTC 301
 DB 241 TTGAGAAAAATGTAATCTACATCAAGGAAAACTTAAAGAAAGCAATTTAAAAAAATTTTC 300
 QY 302 AGATTATACGATTAAGATTTGCAAGAGATGAACAACATGCTTATTTTGGAGGGAG 361

Db 301 AGATTATACCGATAGAGATTGCAAAGAGATGAACCAATGCTTATTTTGGAGGAG 360
Oy 362 TATGTTCAAAAGAACTTAAAGATGATTCGCAATCTTTGAGAAATAGTCTATCAAAAGA 421
Db 361 TATGTTCAAAAGAACTTAAAGATGATTCGCAATCTTTGAGAAATAGTCTATCAAAAGA 420
Oy 422 AACGTGATTAAGTTCCGGAGAGAGTTCTTTTAAGAGCATTCGTGAGCAATCTTAATGAT 481
Db 421 AACGTGATTAAGTTCCGGAGAGAGTTCTTTTAAGAGCATTCGTGAGCAATCTTAATGAT 480
Oy 482 CAGTATATGTAAGAAAAAACTTAAAGATTGGCCGTGCTAGGGGAGGGAAGTGATG 541
Db 481 CAGTATATGTAAGAAAAAACTTAAAGATTGGCCGTGCTAGGGGAGGGAAGTGATG 540
Oy 542 AGTTAACAACTTGTGTCTGAACAGAAAGAGACATGTAGAAATTTTAAATTGAAAAAG 601
Db 541 AGTTAACAACTTGTGTCTGAACAGAAAGAGACATGTAGAAATTTTAAATTGAAAAAG 600
Oy 602 ATAGAGATGCCGTACTCTTAAACAGATGTTTCAGCAGCCTAGAGAGTTTAAAAAG 661
Db 601 ATAGAGATGCCGTACTCTTAAACAGATGTTTCAGCAGCCTAGAGAGTTTAAAAAG 660
Oy 662 AAAACATGCTTGAATTACTCGAACATGCTATTTTACATGGAAATTGCCGAGACGAG 721
Db 661 AAAACATGCTTGAATTACTCGAACATGCTATTTTACATGGAAATTGCCGAGACGAG 720
Oy 722 ATATATATTAATGATATGTAATGGAGGAAATGCCAAGACAAATAATGCTTATATGC 781
Db 721 ATATATATTAATGATATGTAATGGAGGAAATGCCAAGACAAATAATGCTTATATGC 780
Oy 782 CACCGAGAACCCGATTTTGTATCCACTAGGCCAGAGGCTTCAATAGCAGAGATATAGGC 841
Db 781 CACCGAGAACCCGATTTTGTATCCACTAGGCCAGAGGCTTCAATAGCAGAGATATAGGC 840
Oy 842 TGGAGAGATTTAATAAGAGGTAGAGAGGATGGATTTTTATGGAAAGATCATCTTA 901
Db 841 TGGAGAGATTTTATAGAGAGGTAGAGAGGATGGATTTTTATGGAAAGATCATCTTA 900
Oy 902 GAGATGCGACAGCTTTGTGGCATTTGTATCCAAATCTTAAATTTAAAAAGAAAGCG 961
Db 901 GAGATGCGACAGCTTTGTGGCATTTGTATCCAAATCTTAAATTTAAAAAGAAAGCG 960
Oy 962 ACAGAGAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAAATCTCATGAACATGAG 1021
Db 961 ACAGAGAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAAATCTCATGAACATGAG 1020
Oy 1022 CTTTGAAGATTATGTAAGAAAAATGGTTAAGTATGATGGAAGCAAAAAATGGAAG 1081
Db 1021 CTTTGAAGATTATGTAAGAAAAATGGTTAAGTATGATGGAAGCAAAAAATGGAAG 1080
Oy 1082 AATTGCAAAATGATATTAACAAAACTTGCAAAATTTTCACTTCAAAAGTCACTAATAATC 1141
Db 1081 AATTGCAAAATGATATTAACAAAACTTGCAAAATTTTCACTTCAAAAGTCACTAATAATC 1140
Oy 1142 GTCTTTTGTATCAACAAAGGAATATGAAATTTGTGATGGAGAGGCTTCCAAAT 1201
Db 1141 GTCTTTTGTATCAACAAAGGAATATGAAATTTGTGATGGAGAGGCTTCCAAAT 1200
Oy 1202 TTCTTAGCAAGAAAGTTGTGCAAAATGGAGTCCATTTGTTCTATTTTGAAAAAAAT 1261
Db 1201 TTCTTAGCAAGAAAGTTGTGCAAAATGGAGTCCATTTGTTCTATTTTGAAAAAAAT 1260
Oy 1262 GTCCAGATGAGAAAAATGCAATGTAAATATATAGAGCAACATGTTACAAAGAGACTTG 1321
Db 1261 GTCCAGATGAGAAAAATGCAATGTAAATATATAGAGCAACATGTTACAAAGAGACTTG 1320
Oy 1322 ATGCAACGGGCAATTAAGTCTGCAAGAAATATGCGAGGAATGTTACATGCTTCAACA 1381
Db 1321 ATGCAACGGGCAATTAAGTCTGCAAGAAATATGCGAGGAATGTTACATGCTTCAACA 1380
Oy 1382 AAAAGCTGGTTGAAAAAGTTTCAACAAGAAATTAGTAAAGTATGTAGAGAAACGAAAAAG 1441
Db 1381 AAAAGCTGGTTGAAAAAGTTTCAACAAGAAATTAGTAAAGTATGTAGAGAAACGAAAAAG 1440

Oy 1442 AAAACAAAGAAAGTTTCTCAACGATGAATTAATTAATCTGTGTGTACAGCCAGCAAAAG 1501
Db 1441 AAAACAAAGAAAGTTTCTCAACGATGAATTAATTAATCTGTGTGTACAGCCAGCAAAAG 1500
Oy 1502 CAGCCCGGTTGCTTACACATGATCTTGGAAATGAAAAACATCTTTTTCGACAAACACTGG 1561
Db 1501 CAGCCCGGTTGCTTACACATGATCTTGGAAATGAAAAACATCTTTTTCGACAAACACTGG 1560
Oy 1562 ATCAAAAGCAGATTTTCCGACAGATAAAAATTCGAAGAAATTTGGGAGAAAGTGCCAAAG 1621
Db 1561 ATCAAAAGCAGATTTTCCGACAGATAAAAATTCGAAGAAATTTGGGAGAAAGTGCCAAAG 1620
Oy 1622 ATTTAGAGAGGATTCAAAAGAAATTAATGCGCATGTCAATACATGAGCAGCAATGCA 1681
Db 1621 ATTTAGAGAGGATTCAAAAGAAATTAATGCGCATGTCAATACATGAGCAGCAATGCA 1680
Oy 1682 ATGCTTTGGGACATACAGAAATTTTAAAGCAGTTTAAATGGAGTAAACACAAAGATCTT 1741
Db 1681 ATGCTTTGGGACATACAGAAATTTTAAAGCAGTTTAAATGGAGTAAACACAAAGATCTT 1740
Oy 1742 TGAAGAACAAGAAAGTTGTATAAATACCTAAAGAAAAAGTGTAAATGATGCTTAGAA 1801
Db 1741 TGAAGAACAAGAAAGTTGTATAAATACCTAAAGAAAAAGTGTAAATGATGCTTAGAA 1800
Oy 1802 GAGAGATGACCGTTTCTCTTTGTATGTGTCTTCCAAACGCTACGTTGAGCTGATGG 1861
Db 1801 GAGAGATGACCGTTTCTCTTTGTATGTGTCTTCCAAACGCTACGTTGAGCTGATGG 1860
Oy 1862 TAAAGACGTGAAGACAGCGGTGAGATTCAAAAAAATATTAAGGCTCATATATTA 1921
Db 1861 TAAAGACGTGAAGACAGCGGTGAGATTCAAAAAAATATTAAGGCTCATATATTA 1920
Oy 1922 TTGAATTTCTGAATAATATCAATATAAATAATACACACACTGGAAGAAATGTCCCTCTT 1981
Db 1921 TTGAATTTCTGAATAATATCAATATAAATAATACACACACTGGAAGAAATGTCCCTCTT 1980
Oy 1982 GGCATACGTAATGCAATGATTTTCACTAATGTCCAGGTCTTAACGAAAGAAATGATT 2041
Db 1981 GGCATACGTAATGCAATGATTTTCACTAATGTCCAGGTCTTAACGAAAGAAATGATT 2040
Oy 2042 GTACAAAATCAAGAGATGAGAGCCGTTCTTAAGAAAGAGCCTTGGAAGATGCTC 2101
Db 2041 GTACAAAATCAAGAGATGAGAGCCGTTCTTAAGAAAGAGCCTTGGAAGATGCTC 2100
Oy 2102 TCAAGTAGAGCTTCAAGAAAAATTTGACTGAATCTAATGTGAACCTGCAATTGAAA 2161
Db 2101 TCAAGTAGAGCTTCAAGAAAAATTTGACTGAATCTAATGTGAACCTGCAATTGAAA 2160
Oy 2162 GATATTTGTACGTAACGAGGAAACGTAAATAATGCGTCAATCAGTGGCTTATGCAAGCTA 2221
Db 2161 GATATTTGTACGTAACGAGGAAACGTAAATAATGCGTCAATCAGTGGCTTATGCAAGCTA 2220
Oy 2222 ACACCAAGGATTAATCTCGAAGAAAGATGAGATGCTTAGAAAGAAATCTGTGAGAAAT 2281
Db 2221 ACACCAAGGATTAATCTCGAAGAAAGATGAGATGCTTAGAAAGAAATCTGTGAGAAAT 2280
Oy 2282 TAGTGAAGAAAGTGAAGAACAGTGCAAAGCATTAACCAACAGAAATTAGAGCAACCGGAG 2341
Db 2281 TAGTGAAGAAAGTGAAGAACAGTGCAAAGCATTAACCAACAGAAATTAGAGCAACCGGAG 2340
Oy 2342 CTGATCTTAAAAAAGATTATAGACATATGAGAACTTAAAGAAACGTGCAAGAGAACAA 2401
Db 2341 CTGATCTTAAAAAAGATTATAGACATATGAGAACTTAAAGAAACGTGCAAGAGAACAA 2400
Oy 2402 TGAACAAGTCCAGTCTTGTGTCACATTAAGAAAAAGAAAGTAAATGATCAAAAA 2461
Db 2401 TGAACAAGTCCAGTCTTGTGTCACATTAAGAAAAAGAAAGTAAATGATCAAAAA 2460
Oy 2462 GTAATAGCAAAAAACAAGATAGATGCGTTTCAACGAGACTTCAAGATACCAAAAAAC 2521
Db 2461 GTAATAGCAAAAAACAAGATAGATGCGTTTCAACGAGACTTCAAGATACCAAAAAAC 2520

Qy	2522	ATGTGAATACTACGAGAGGAGTAAAGATGATTCGGTAAACAGAAATTAGAACTTAAAG	2581
Db	2521	ATGTGAATACTACGAGAGGAGTAAAGATGATTCGGTAAACAGAAATTAGAACTTAAAG	2580
Qy	2582	CATTGATTTGGCAGCAGAAAGTATTGGAAAGATATGTAGATTTGAGAGAAAGATGATA	2641
Db	2581	CATTGATTTGGCAGCAGAAAGTATTGGAAAGATATGTAGATTTGAGAGAAAGATGATA	2640
Qy	2642	AATTGGAATCAGATTCGAGAAATTAAGAGAGTTTCGAAAGACTTGAAAGAGTATGCAAAA	2701
Db	2641	AATTGGAATCAGATTCGAGAAATTAAGAGAGTTTCGAAAGACTTGAAAGAGTATGCAAAA	2700
Qy	2702	AGATTAAATTAAGGCTTTCGCAATCGAATCGCTTGAAGGTAGACCGCAGCAACAGTGA	2761
Db	2701	AGATTAAATTAAGGCTTTCGCAATCGAATCGCTTGAAGGTAGACCGCAGCAACAGTGA	2760
Qy	2762	CAGAAAGTACACGACAACTACAAACAACAAACAAACCGTTGCCGATCCGAAAGCAACGG	2821
Db	2761	CAGAAAGTACACGACAACTACAAACAACAAACAAACCGTTGCCGATCCGAAAGCAACGG	2820
Qy	2822	AATGCAAAATCTTTACAGACACACACATGGGTTTACACAGACATTCGACACACAAAGCA	2881
Db	2821	AATGCAAAATCTTTACAGACACACACATGGGTTTACACAGACATTCGACACACAAAGCA	2880
Qy	2882	CGTCTACATCAGATCTACATCATCAATCAAAATAATTAACATTCGACACACAGGCGATGCA	2941
Db	2881	CGTCTACATCAGATCTACATCATCAATCAAAATAATTAACATTCGACACACAGGCGATGCA	2940
Qy	2942	AACCAACCAAGTGTACGACAGAGAGAGAGATGATGACAGAGACGTGAAACCGAGTAGG	3001
Db	2941	AACCAACCAAGTGTACGACAGAGAGAGATGATGACAGAGACGTGAAACCGAGTAGG	3000
Qy	3002	GGCTGAGATGAGTGGGTGGAATGTGATGAGGGGGGTGATGTGCAATGTTATTTTCGT	3061
Db	3001	GGCTGAGATGAGTGGGTGGAATGTGATGAGGGGGGTGATGTGCAATGTTATTTTCGT	3060
Qy	3062	TCATGATTTAG 3072	
Db	3061	TCATGATTTAG 3071	

RESULT 3	AF033210	3056 bp	DNA	linear	PLN 10-SBP-1998
LOCUS	Pneumocystis carinii f. sp. hominis clone HUMSG33 major surface glycoprotein (MSG) gene, partial cds.				
DEFINITION	AF033210				
ACCESSION	AF033210.1	GI:3560516			
VERSION					
KEYWORDS	Pneumocystis jirovecii				
SOURCE	Pneumocystis jirovecii				
ORGANISM	Pneumocystis jirovecii				
REFERENCE	1 (bases 1 to 3056)				
AUTHORS	Garbe, T.R. and Stringer, J.R.				
TITLE	Molecular characterization of clustered variants of genes encoding major surface antigens of human Pneumocystis carinii				
JOURNAL	Infect. Immun. 62 (8), 3092-3101 (1994)				
MEDLINE	94314421				
PUBMED	7518806				
REFERENCE	2 (bases 1 to 3056)				
AUTHORS	Mel, O., Turner, R.E., Sorial, V., Kilmington, D., Angus, C.W. and Kovacs, J.A.				
TITLE	Characterization of major surface glycoprotein genes of human Pneumocystis carinii and high-level expression of a conserved region				
JOURNAL	Infect. Immun. 66 (9), 4268-4273 (1998)				
MEDLINE	98380374				
PUBMED	9712777				
REFERENCE	3 (bases 1 to 3056)				
AUTHORS	Mel, O., Turner, R., Sorial, V., Kilmington, D., Angus, C.W. and Kovacs, J.A.				
TITLE	Direct Submission				

JOURNAL	Submitted (07-NOV-1997) Critical Care Medicine Dept., National Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD 20892-1662, USA
FEATURES	Location/Qualifiers
source	1. .3056
	/organism="Pneumocystis jirovecii"
	/mol_type="genomic DNA"
	/specific_host="Homo sapiens"
	/sub_species="hominis"
	/db_xref="taxon:42068"
	/clone="HUMSG33"
gene	/note="derived from HIV-infected human with P. carinii pneumonia"
	<1. .3056
	/gene="MSG"
CDS	<1. .3056
	/gene="MSG"
	/note="surface antigen"
	/codon_start=3
	/product="major surface glycoprotein"
	/protein_id="AAC34973.1"
	/db_xref="GI:3560517"
	/translation="AAVKKQVTVASQYDDEVNIALIILQEDMEDTKKSLKELK EELKASLDMKHYKMLKDFCGNGKASKATKCOGLQAKYTGCTNFKTKGLPALTN PSDDNCKESBRQCLFLGACHNVEDCNKRLNLCYOKRGDVAEVLIALRSDPLNT ETHEKKKEI CPTLORESNELTDLCLNOKTKCNIIKEKDKCTTLKANVATLGSFK KEICBELBCYFVYIGNCGDDPIIKCELEGKOEONIVYIPGPDDEPRPPTALAE DIDULBYKAEDEGVFIGKHLRLDATALTLVKKDDTKRNNIGSKKKILDEKCK NSOHEALKNLKNNSNPNAYGKERCELEDDIKTKTNLPTLLKNLTPNDKIVEM RKLPTFLINEDCARLBSYCFYERKACPNABECBNILAAACYKGLDGRANKVLOENR GLRGSNOSWLFKEFOGLVYCKELKENKGSFNDIEFLVCPAPAKARLLTHDQWR VIFLRQOLDQKRDPTDCKEIGKKKODIGKSKELTWPCHLLEOQCNLGTTEILK QVILDEKDTLKHQESCVKYLKCKNKMRSRGDRSFVCFVONATELMLKVKYDKC EEFKRNKASYLIEFLNNNTNKITTLERNCPWHITCNRPSPNGPGLTKENSCRTIK HCEPFYKRALDELALKVELQKLTDSKCEPALRCCTVAGVNNVASISGLCKANTD NSGKSDBARKELECKLVKEVEQCKALPTELQPADLKDXYTELKVRLEBAN KSLVLSLTKENSNVSKSNKNKDKNVNGLODTTKHYKILRGVQDVSVLELAK AFDLAEEVPRGVYDIKERCNKLRSDDRICEKDCLEVKKINACNLPLEKPHF TVTESITTTTTTTTTTADPKATRECKSLQTTDTVTCSTHTSTSTSTSTSTSTLTS TRCKPPTCTTGDAEDVPSBGLRVSGMVMKGVIAVVISPMI"

ORIGIN

Query Match	75.5%	Score 2320.8	DB 8	Length 3056
Best Local Similarity	66.0%	Pred. 0		
Matches 2646	Conservative 0	Mismatches 404	Indels 27	Gaps 6
Qy	2	TGGCGCGGGCGGTCAAGCGGCAGG--CAGCAGGACACAGAAATGAGCATGATGAGAGAAC	58	
Db	1	TGGCGCGGGCGGTCAAGCGGCAGGTAACAGAGCATCAGGCGCAATATGATGATGAAATGA	60	
Qy	59	ATGTTTAACTTTAATTTCTAAAGAAATGAGCTTAATGAGCAGGAATGCAAAAAAATAC	118	
Db	61	ATATTTTGGGGTGAATTTCTACAGAAAGATCAATGAGAAATACAAATGCAAAAAAGTT	120	
Qy	119	TAAAAAATAATTCGCAAGAAATGACTGAGCAAAATCAATTAATATGAAACAATACAGAA	178	
Db	121	TAGAAAAAATCTCGGAGAGATTTGAAAAAAGCATCTAGACATGAAAAAATGACTAAAA	180	
Qy	179	AACCTAAAGGTTTTTTCGAAAGATGAAAGACAGATACAAATGCAAGAACTGAAAGCCA	238	
Db	181	TGCTTAAAGATTTCTGTGAAATGGAAGCAAGTAAGCAAAATATGTCAAGGTC	240	
Qy	239	ATATTGAAAAAATGATCTACATCAAGAAAACTTAAAGAACTTAAAAAATAAAAA	298	
Db	241	TACAAAGCAAAAGTTACGGGAAATGTCAAAATTTTAAACACAAACCTAGAGACACGCT	300	
Qy	299	TTGAGATTATACGATTAAGATTTGCAAGAGATGAAAGATGAAACAATGCTATTTTGGAG	358	
Db	301	TAACAATTCATCAGATGATTAATTTGCAAGAGATGAAACAATGCTATTTTGGAGG	360	
Qy	359	GAGTATGTTCAAAAGAACTTAAGATGATTTGCAATCTTTGAGAAATAGTGCTATCAAA	418	
Db	361	GAGCATG---CAATTAATCTGTAGAAAGATTTGTAAACAAACTTAAGGAATCTATGTACAGA	417	

419 AGAAAGTGTAAAGTTGGGAGAGATTCTTTAAGACACTTGTGACATCTTAATG 478
418 AAAACGTGACGGAGTAGAGAAAGTCTTTTGAAGGCACTTGATGATCTCAATA 477
479 GATCGTATATGTAAGAAAAAACTTAAAGATTTGCCCTGTCATGGGGAGGAAAGT 538
478 AAACGAAACACATAAAAAACTGAAGAGATTTGCCAGTCTTGACAGGGGAAAGTA 537
539 ATGAGTTAAACAACCTTGCTGTGACACAGAAAGACATGTAAAGATTTTAAATGAAA 598
538 ATGATTTAAACGACTTGCTGTTGAACAGAAAAACGTGGAGAAATTTAAAGAAA 597
599 AAGATTAAGAGTGGGTACTTTAAAACAGATGTTTACAGCAGCTAGAGAGTTTAAA 658
598 AAGATTAAGAAAAAGCACTCTTTAAAGCAATGTTGCAACAGCACTTGAAGTTTAAA 657
659 AAGAAACATGTTCTTGAATTAACGCAACATGCTATTTTACATGGAAATTTCCGAGACG 718
658 AAGAAATATGCTTGAATTAACGCAACATGCTATTTTACATGGAAATTTCCGAGACG 717
719 ACGATATATTAATTAATGTAATGGAATGGAGCAAAATGCAAGAACAAATATGCTTATA 778
718 ACGATATATTAATTAATGTAATGGAATGGAGCAAAATGCAAGAACAAATATGCTTATA 777
779 TGCACACGAGACCCGATTTTGAATCCAACTAGGCCAGAGCTACAAATAGCAGAGATAG 838
778 TACCAACAGAGACCCGATTTTGAATCCAACTAGGCCAGAGCTACAAATAGCAGAGATAG 837
839 GGGCTGAAGAGTTTAAAGAGATTAAGAGAGATGAGAGTTTAAATGGAAAGATTCAT 898
838 ACCTGATGAGCTTTTAAAGAGATTAAGAGAGATGAGAGTTTAAATGGAAAGATTCAT 897
899 TAAGAGATGCGACAGCTTGTGTGGCAATTTG---ATCCAGATTCATGCTTAAAAAA 955
898 TAAGAGATGCGACAGCTTGTGTGGCAATTTG---ATCCAGATTCATGCTTAAAAAA 957
956 AAGACGACAAAGAGAAATGCGAAGAACCTTCAAAAAAGCTGCAAAAATCTCATGAC 1015
958 ATAAATGCGAGAGAAATGCAATTAAGATTCGAAAGATTAATGCAAAAACTCTCAACGC 1017
1016 ATGAGGCTTTAGAAAGTTATGTAAGAAAAATGTTTAAAGTAAGTGAAGCAAGAAAAAT 1075
1018 ATGAAGCTCTTAAAAATTTATGTAATTAATTAATGATCTTAATGCAATGAAAAAANAAT 1077
1076 GTGAAGAAATGCAAAATGATATTAACAACCTTGCAAAAATTTTCACTTCAAAAGTCAGTA 1135
1078 GCAAGAAATTAAGAGATATTTAAAAAACAATGCACAAACCTCAACCAAGATTTCTTA 1137
1136 ATAAATGCTTTTGAATCCACAAAGAAATTAAGAAATGTTGAGATGGAGAGGTTGC 1195
1138 AAAACCATCTTATATATCC-----AAATGATTAATTTGTGAGTGAAGAAAACTGC 1188
1136 CAACATTTCTTGAACAGAGATGTCGCAAAATTTGAGTCTAATGTTTCTTATTTTGA 1255
1189 CAACATTTCTTGAACAGAGATGTCGCAAAATTTGAGTCTAATGTTTCTTATCTAGAAA 1248
1256 AAAAAATGTCAGATGAGAGAAATGCAATGTAATAATTAAGACCAATGTTCAAAAGAG 1315
1249 AAGCTTGTCAATGCAAGAGAGAGAGATGTAATTTGAGGCGACGTTTATTAAGAGAG 1308
1316 GACTTGATGACGCGCAATTAAGTGTGCAAGAAATTAATGCGAGAAATGTTTACATGTT 1375
1309 GCGTTGATGAGCGGCAATTAAGTGTGCAAGAAATTAATGCGTGGTATTAATGCTGTT 1368
1376 CAACAAAGAGTGGCTGTAAGAGTTTCAACAAGATTTGTAAGATTAAGATGAGAAACGTA 1435
1369 CAACATCAAGATTTGCTTAAGAGTTTCAACAAGATTAAGATTAAGATTAAGAGAGCT-- 1426
1436 AAAAAAGAAACAAAGAGATTTCTCAACGATGAATTTATTTCTGTTGTACAGCCAG 1495
1427 -AAAAAGAAATTAAGAGAGTTTCCCAACGATGAATATTTGTTCTGTTGTACAGCCAG 1485

1496 CAAAGACAGCCCGTGTGCTTACATATGATCTTCAATGAAAACTATCTTTTACGACAC 1555
1486 CAAAAGCTGCAAGATTTCTTACACAGATCACTAAATGAAGGTTATCTTTTACGACAC 1545
1556 AACTGATTCAAAAGACGAGATTTCCGACAGATTAATAATGCAAGAAATGGGAGAAAGT 1615
1546 AATTGATTCAAAAGAGAGATTTCCGACAGATTAAGATGCAAGAAATTAAGGAAAAAT 1605
1616 GCCAAGTTTAAAGAGAGATTCAAAAGAAATTAATGATGCGATGATCACTGAGACAGC 1675
1606 GCCAAGATTTTAAAGAGATTCAAAAGAAATTAATGATGCGATGATCACTGAGACAGC 1665
1676 AATGCAATCGTTGGGACCTACAGAAATTTTAAAGCAGGTTTATGGAATGAACACAAAG 1735
1666 AATGCAATCGCTTGGGACCTACAGAAATTTTAAAGCAGGTTTATGGAATGAACACAAAG 1725
1736 ATACTTGAAGAACCAAGAAAGTTGTGTAAATACCTAAAGAAAGATGATTAATAGT 1795
1726 ATACTTGAAGAACCAAGAAAGTTGTGTAAATACCTAAAGAAAGATGATTAATAGT 1785
1796 CTAGAAGAGAGATGACCGTTTCTTTTGTATGTCTTCAAAAACGCTACGTTGAGC 1855
1786 CTAGAAGAGAGATGACCGTTTCTTTTGTATGTCTTCAAAAACGCTACGTTGAGC 1845
1856 TGATGCTTAAAGACGTAAGAGACAGCTGTGAACTTCAAAAATTAATTAAGCTTCAT 1915
1846 TGATGCTTAAAGACGTAAGAGACAGCTGTGAACTTCAAAAATTAATTAAGCTTCAT 1905
1916 ATATTATGAATTTCTTGAATAATATCAATTAATAATTAACAACCTGAAAGAAATGTC 1975
1906 ATATTATGAATTTCTTGAATAATATCAATTAATAATTAACAACCTGAAAGAAATGTC 1965
1976 CCTTGGCATACTGATTAATGATTAATTTTCACTAATTTGTCAGGCTTTACGAAGAGA 2035
1966 CCTTGGCATACTGATTAATGATTAATTTTCACTAATTTGTCAGGCTTTACGAAGAGA 2025
2036 ATAGTGTACAAAATCAAGAGATGTCGAGCCGTTCTAATAAGAAAGCCTTGGAG 2095
2026 ATAGTGTACAAAATCAAGAGATGTCGAGCCGTTCTAATAAGAAAGCCTTGGAG 2085
2096 ATGCTCTCAAGATGAGGCTTCAAGAAAAATGACTGAATTAATCTTAATGTAACCTGCAT 2155
2086 ATGCTCTCAAGATGAGGCTTCAAGAAAAATGACTGAATTAATCTTAATGTAACCTGCAT 2145
2156 TGAAGATATTTGTAACGTAACGCGGAAAACGTAATTAATGCTCAATCAGTGGCTTAAGCA 2215
2146 TGAAGATATTTGTAACGTAACGCGGAAAACGTAATTAATGCTCAATCAGTGGCTTAAGCA 2205
2216 AAGCTAACCAAGATTAATCTGGAAGAGATGTAAGAGATGTAAGAAAGAACTGCTG 2275
2206 AAGCTAACCAAGATTAATCTGGAAGAGATGTAAGAGATGTAAGAAAGAACTGCTG 2265
2276 AGAAATTAAGAAAGATGGAAGAACAGTSCAAGCACTTACCAACAGAAATTAAGCAAC 2335
2266 AGAAATTAAGAAAGATGGAAGAACAGTSCAAGCACTTACCAACAGAAATTAAGCAAC 2325
2336 CGGACGCTGATCTAAAAAAGATTAATTAAGACATATGAGAACTTAAGAAACGTCGAGAG 2395
2326 CGGACGCTGATCTAAAAAAGATTAATTAAGACATATGAGAACTTAAGAAACGTCGAGAG 2385
2396 AAGCAATGAACAATGTCAGCTTGTGTTGTCACTTAATAAGAAAAACAAAGTAAATGAT 2455
2386 AAGCAATGAACAATGTCAGCTTGTGTTGTCACTTAATAAGAAAAACAAAGTAAATGAT 2445
2456 CAAAAGATTAAGCAAAACAAAGATTAAGATGTCGTTCAACGGAATTAAGATTAAGCA 2515
2446 CAAAAGATTAAGCAAAACAAAGATTAAGATGTCGTTCAACGGAATTAAGATTAAGCA 2505
2516 CAAAACATGTGAAAAATCTACGAGAGGAGTTAAGATGATCCGTAACAGAAATTAAGAG 2575
2506 CAAAACATGTGAAAAATCTACGAGAGGAGTTAAGATGATCCGTAACAGAAATTAAGAG 2565
2576 CTAAGACATTTGATTTGGCAGACAGAAATTTTGAAGATATGATGATTTGAAGAAAGAT 2635

|||||
Db CTAAGCATTTGATTTGGCAGCAGAGATTTGGAAGATATGATTTGGAAGAAAGAT 2625
Qy 2636 GTATTAATTTGATTCAGATTTGCAATTTAAGAGATTTGCAAGACTTGAAGAAAT 2635
Db 2626 GTATTAATTTGATTCAGATTTGCAATTTAAGAGATTTGCAAGACTTGAAGAAAT 2685
Qy 2696 GCAAAAAGATTTAATGAAGCTTGTGCAATCTGAAGCCCTGTGAGATGAGCCGACGMA 2755
Db 2686 GCAAAAAGATTTAATGAAGCTTGTGCAATCTGAAGCCCTGTGAGATGAGCCGACGMA 2745
Qy 2756 CAGTGACAGAAAGTACACGACAACTACACACACACACACACACACACACACAC 2815
Db 2746 CAGTGACAGAAAGTACACACGACAACTACACACACACACACACACACACACAC 2805
Qy 2816 CACGGAATGCAATCTCTTACACACACACACACACACACACACACACACACACAC 2875
Db 2806 CACGGAATGCAATCTCTTACACACACACACACACACACACACACACACACACAC 2865
Qy 2876 CACGCAAGTCTACTATCATCATCTACATCAATCAATCAATCAATCAATCAATCA 2935
Db 2866 CACGCAAGTCTACTATCATCATCTACATCAATCAATCAATCAATCAATCAATCA 2925
Qy 2936 GATGCAAC 2995
Db 2926 GATGCAAC 2979
Qy 2996 GTGAGGGGCTGAGATGATGCTGGTGAATGTGATGAGGGGGGTATGATGCAAT 3055
Db 2980 GTGAGGGGCTGAGATGATGCTGGTGAATGTGATGAGGGGGGTATGATGCAAT 3039
Qy 3056 TTTCGTCATGATTTAG 3072
Db 3040 TTTCGTCATGATTTAG 3056

RESULT 4
AR438439
LOCUS AR438439 3054 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 11 from patent US 6664053.
ACCESSION AR438439
VERSION AR438439.1 GI:42663294
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3054)
AUTHORS Kovacs,J.A., Huang,S., Maehr,H., Fischer,S.H., Gyll,V.J. and Mei,Q.
TITLE Identification of a region of the major surface glycoprotein (MSG)
JOURNAL gene of human Pneumocystis carinii
FEATURES
SOURCE Location/Qualifiers
1..3054
/organism="unknown"
/mol_type="genomic DNA"

Query Match 75.5%; Score 2318.8; DB 6; Length 3054;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 2644; Conservative 0; Mismatches 404; Indels 27; Gaps 6;

Qy 181 CTTAAAGTTTTGCGAAGATGAAAAAGCAGATACAAATATGCAAGAACTGAAAGCCAA 240
Db 181 CTTAAAGATTTCTGTGAAATGGAAGCAAGTAAACCAATATCAAAATGCAAGAACTG 240
Qy 241 ATTGAGAAAAATGTAATCAATCAAGAAAACTTAAAGAGCAATTTAAAAAAATTT 300
Db 241 CAAAGCCAAAGTTACGGGAAATGTAACAAATTTTAAAAACAAAAAGCTAGACCGCT 300
Qy 301 CAGTTTAAACGATTAAGATTTGCAAGAGATGAAACAAATGCTATTTTGGAGGA 360
Db 301 ACAATTCATCAGATTAATTTGCAAGAGATGAAACAAATGCTATTTTGGAGGA 360
Qy 361 GTATGTTCAAAAGAACTTAAAGATGATTCATATCTTTGAAATTAAGTCTATCAAA 420
Db 361 GCATG---CCATATCTTTGAAAGATTTTAAACAAATTAAGAAATCTATGTTACAGA 417
Qy 421 AAACGTATTAAGTTGCGGAAAGATTTTAAAGACACTTGTAGCATTTAATGGA 480
Db 418 AAACGTATTAAGTTGCGGAAAGATTTTAAAGACACTTGTAGCATTTAATGGA 477
Qy 481 TCAGTCATATGTAAGAAAAAATTAAGAGATTTGCCCTGCAATGGGAGGAAATGAT 540
Db 478 ACAGAAACACATGAAAAAATTAAGAGATTTGCCCTGCAATGGGAGGAAATGAT 537
Qy 541 GAGTTTAAACAACTTGTCTGAAACAGAAAGACATGTAAGATATTTAATTTGAAAA 600
Db 538 GAATTTAAGCACTTGTCTTGAACAGAAAGACATGTAAGATATTTAATTTGAAAA 597
Qy 601 GATTAAGAGTGGGATCTTTTAAACAGATTTTACAGACACTAGGAAGTTTAAAA 660
Db 598 GATTAAGAGTGGGATCTTTTAAACAGATTTTACAGACACTAGGAAGTTTAAAA 657
Qy 661 GAAACATGCTGAAATTAATCTGAAACATGCTATTTTAACTGGAAATTCGGAGAGAC 720
Db 658 GAAATATGCTGTAATTAATCTGAAACATGCTATTTTAACTGGAAATTCGGAGAGAC 717
Qy 721 GATTAATTAATGATTAATGAGATTTGGAGAAATGCAAGAACTATTTGCTTATATG 780
Db 718 GATTAATTAATGATTAATGAGATTTGGAGAAATGCAAGAACTATTTGCTTATAT 777
Qy 781 CCACAGAGCCCGATTTTGAATCCATGAGCCAGAGGCTTACATGACAGAGATATG 840
Db 778 CCACAGAGCCCGATTTTGAATCCATGAGCCAGAGGCTTACATGACAGAGATATG 837
Qy 841 CTGGAAGATTTTAAAGAGATTTAGAGAGATTTTAAAGAGATTTTAAAGAGAT 900
Db 838 CTGGAAGATTTTAAAGAGATTTAGAGAGATTTTAAAGAGATTTTAAAGAGAT 897
Qy 901 AGAGATGCGACAGCTTTGTCATTTGTC---ATCCAGATTTCTAGTCTTAAAAA 957
Db 898 AGAGATGCGACAGCTTTTATGACCTTTGTCATTTGTCATTTGTCATTTGTCAT 957
Qy 958 GACGACAAAGAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAAATCTCATGA 1017
Db 958 AATATCGAGAAAAATGCAATTAAGATTTCTGAGATTAAGCAAAAAATCTCATGA 1017
Qy 1018 GAGCTTTAAGAAATTTATGTAAGAAATTTGTTAATGATGATGTAAGCAAAAAAT 1077
Db 1018 GAACTCTTAAAAATTTATGTAAGAAATTTGTTAATGATGATGTAAGCAAAAAAT 1077
Qy 1078 GAAAGATTTGCAAAATGATTTAACAATTTTCAAAATTTTCACTTCAAAAGTCA 1137
Db 1078 AAAGATTTAAGAAAGATTTAACAATTTTCAAAATTTTCACTTCAAAAGTCA 1137
Qy 1138 AATGCTCTTTTATGTCACAAAGAAATTAAGAAATTTGTTGATGGAAGGATGCCA 1197
Db 1138 AACATCTTTATGATTC-----AATGATTAATTTTGTGATGGAAGGATGCCA 1188
Qy 1198 ACATTTCTTACAGAAAGATTTGCGAATTTGGAAGTCTATTTGTTTGA 1257
Db 1189 ACATTTCTTACAGAAAGATTTGCGAAGATTTGGAAGTCTATTTGTTTGA 1248
Qy 1258 AATGTCAGATGAGAAATGCAATGTAATAATTAAGCAACATGTTTCAAAAGAGA 1317

Db	1249	GCTTGTCCAAATGCCAAAGAGAGTGTAATGAATTTGAGGGCAGCGCTGTATTAAGAGAGGG	1308
Oy	1318	CTTGATGCA CGGGCAATTAAGTCTGCAAGAAATATGCGAGAGTGTACATGCTTCA	1377
Db	1309	CTTGATGACGGGCAAAATAAAGTCTGCGAAGAAATAATGCGGGGTATTTACGTGTTCA	1368
Oy	1378	AACAAACCTGGCTTGAAGTTTCAACAAAGATTAGTAAAGTATGAGAAACTGAA	1437
Db	1369	AATCAAGTTGGCTTAAAGAGTTTACAAAGATTAGTAAAGTATGTAAGAGCT--A	1428
Oy	1438	AAAGAAACAAAGGAAGTTTCTCAAAACGATGATATTAATTTCTGTGTGTCAGCCAGCA	1497
Db	1426	AAAGAAATTAAGGAAGTTTCCCAACGATGAAATATTTGTTCTGTGTGTCAGCCAGCA	1485
Oy	1498	AAAGCAGCCCGGTGCTTACACATGATCTTGCAATGAAACCTATCTTTTACGACAA	1557
Db	1486	AAAGTCGACGATTACTTACACAGATCATCAATGAGGGGTATCTTTTACGACAA	1545
Oy	1558	CTGATTCAAAACCGAGATTTCCCGCAAGTAAATTTGCAAGGAATTTGGGAGAAAGTGC	1617
Db	1546	TTGGATCAAAAAGAGAGATTTTCCGACAGATTAAGCTCAGAGGAATTTGGGAAAAATGTC	1605
Oy	1618	CAAGATTTAGGAGAGAGATTCAAAAGAAATTTACATGGCCATGTCATACCTGAGCAGCA	1677
Db	1606	CAAGATTTAGGAGAGAGATTCAAAAGAAATTTACATGGCCATGTCATACCTGAGCAGCA	1665
Oy	1678	TGCAATCGCTTGGGGACTACAGAAATTTTAAAGCAGTTTTATTTGGATGAACACAAAGAT	1737
Db	1666	TGCAATCGCTTGGGGACTACAGAAATTTTAAAGCAGTTTTATTTGGATGAACACAAAGAT	1725
Oy	1738	ACTTTGAAGAAGCCAGAAAGTTGTGTAAATATCTTAAAGAAAGTGTAAATTAATGTGCT	1797
Db	1726	ACTTTGAAGAAGCCAGAAAGTTGTGTAAATATCTTAAAGAAAGTGTAAATTAATGTGCT	1785
Oy	1798	AGAAGAGAGATGACCGTTTCTCTTTGTATGTGCTTCAAAAAGCTACGCTGAGCTG	1857
Db	1786	AGAAGAGAGATGACCGTTTCTCTTTGTATGTGCTTCAAAAAGCTACGCTGAGCTG	1845
Oy	1858	ATGTGTAAGAAGCGTGAAGAAGCAGGTGTGAAGTATCAAAAAAATATTAAGCTTCATAT	1917
Db	1846	ATGTGTAAGAAGCGTGAAGAAGCAGGTGTGAAGTATCAAAAAAATATTAAGCTTCATAT	1905
Oy	1918	ATTATTTGAATTTCTTGAATAATATCAAAATTAATTAACAACCTGGAAGAAATGTGCC	1977
Db	1906	ATTATTTGAATTTCTTGAATAATATCAAAATTAATTAACAACCTGGAAGAAATGTGCC	1965
Oy	1978	TCTTGGCATAAGTATTTGCAATGATTTTCACTATTTGCCAGGCTTACGAAAGAGAT	2037
Db	1966	TCTTGGCATAAGTATTTGCAATGATTTTCACTATTTGCCAGGCTTACGAAAGAGAT	2025
Oy	2038	AGTTGTACAAAATATCAGAGAGCATGTGAGCCGTTCTATTAAGAAGAAAGCCTTGGAGAT	2097
Db	2026	AGTTGTACAAAATATCAGAGAGCATGTGAGCCGTTCTATTAAGAAGAAAGCCTTGGAGAT	2085
Oy	2098	GCTCTCAAGTAGAGCTTCAAGGAAATTTGACTGATTAATCTAAATGTGAACCTGCATTTG	2157
Db	2086	GCTCTCAAGTAGAGCTTCAAGGAAATTTGACTGATTAATCTAAATGTGAACCTGCATTTG	2145
Oy	2158	AAAAGATATTTGTACAGTAGACGGGAAACGTAAATATTTGGTCAATCAGTGGCTTATGCAA	2217
Db	2146	AAAAGATATTTGTACAGTAGACGGGAAACGTAAATATTTGGTCAATCAGTGGCTTATGCAA	2205
Oy	2218	GCTAACACCAAGATTAACCTTGGAAGAAGTGTAGAGATGCTAGAAAGAACTCTGTAG	2277
Db	2206	GCTAACACCAAGATTAACCTTGGAAGAAGTGTAGAGATGCTAGAAAGAACTCTGTAG	2265
Oy	2278	AAATTTAGTGAAGAGTGGAGAAACAGTGCACAAAGCTTACCAACGAATTTAGACAAACCG	2337
Db	2266	AAATTTAGTGAAGAGTGGAGAAACAGTGCACAAAGCTTACCAACGAATTTAGACAAACCG	2325
Oy	2338	GCAGCTGATCTTAAAAAAGATTTATAGACATATAGGAACCTTAAGAAACGTGCAGAGAA	2397

Db		2326	GCACGTGATTTAAAAAAGATTATAAGACAATATGAGGAACCTTAAGAACAACGTGCAGAGAA	2385
OY		2398	GCAATGAACAAGTCCACTCTTTGTCTCACTCATTTAAGAAAAACGAAGTAATGTATCA	2457
Db		2386	GCAATGAACAAGTCACAGCTCTGTTTTGTCACTCATTTAAGAAAAACGAAGTAATGTATCA	2445
OY		2458	AAAAGTATATGCAAAAAACAAGATPAGAATGCCCTTCAAACGGACTTCAAGATPACCACA	2517
Db		2446	AAAAGTATATGCAAAAAACMAGATPAGAATGCCCTTCAAACGGACTTCAAGATPACCACA	2505
OY		2518	AAACATGTGAAAAATATCTACGAGAGGGAGTTAAGGATGTATCCGTACAGATTTAGAACT	2577
Db		2506	AAACATGTGAAAAATATCTACGAGAGGGAGTTAAGGATGTATCCGTACAGATTTAGAACT	2565
OY		2578	AAACATTTTGATTTTGGCAGCAGAAAGTATTTTGAAGATATGTAGTATTTGAAGAAAGTGT	2637
Db		2566	AAACATTTTGATTTTGGCAGCAGAAAGTATTTTGAAGATATGTAGTATTTGAAGAAAGTGT	2625
OY		2638	AATTAATTTGGAATCAGATTGCGAAGTTAAGAGAGATTGCAAGAAGCTTAGAAGATGTATGC	2697
Db		2626	AATTAATTTGGAATCAGATTGCGAAGTTAAGAGAGATTGCAAGAAGCTTAGAAGATGTATGC	2685
OY		2698	AAAAAGATTAATAAGCTTGTCGCNATCTGAGCCTCTGGAGGTGAAGCCGCACGAACA	2757
Db		2686	AAAAAGATTAATAAGCTTGTCGCNATCTGAGCCTCTGGAGGTGAAGCCGCACGAACA	2745
OY		2758	GTGACAGAAAGTACAACGACAACCTACTACAACAACAACCGTTGCCGATCCGAAGCA	2817
Db		2746	GTGACAGAAAGTACAACGACAACCTACTACAACAACAACCGTTGCCGATCCGAAGCA	2805
OY		2818	ACGGAATGCAAAATCCTTACAGACAACGACAACATGGGTTACACGACATGACACACA	2877
Db		2806	ACGGAATGCAAAATCCTTACAGACAACGACAACATGGGTTACACGACATGACACACA	2865
OY		2878	AGCACGTCTACTATCACATCTTACCATCACATCAAAATTAACATTGACATCAACGAGCGA	2937
Db		2866	AGCACGTCTACTATCACATCTTACCATCACATCAAAATTAACATTGACATCAACGAGCGA	2925
OY		2938	TGCAAACCAACCAAGTGTAACGACAGAGAGAGAGATGATGACGAGACGTGAACCGAGT	2997
Db		2926	TGCAAACCAACCAAGTGTAACGACAG-----GGATGATCAGAAAGCGTGAAACCCAAGT	2979
OY		2998	GAGGGCGCTGAGGATGAGATGGGTGGAATGTGATGAGGGGGGTGATAGTACCAATCGTTATT	3057
Db		2980	GAAAGCTTTGAGGGTGAGACCGGGTGGAAATGTATGAGGGGGGTGATAGTACCAATCGTTATT	3039
OY		3058	TCGTTTCATGATTTAG 3072 	
Db		3040	TCGTTTCATGATTTAG 3054	
RESULT 5				
LOCUS	AR438437		3084 bp	DNA
DEFINITION	Sequence 7 from patent US 6664053.		Linear	PAT 20-FEB-2004
ACCESSION	AR438437			
VERSION	AR438437.1		GI:42663292	
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unclassified.			
REFERENCE	1 (bases 1 to 3084)			
AUTHORS	Kovacs,J.A., Huang,S., Masur,H., Flecher,S.H., Gill,V.J. and Mei,Q.O.			
TITLE	Identification of a region of the major surface glycoprotein (MSG) gene of human Pneumocystis carinii			
JOURNAL	Patent: US 6664053-A 7-16-DEC-2003;			
FEATURES	Location/Qualifiers			
source	1..3084			
ORIGIN	/organism="unknown"			
	/mol_type="genomic DNA"			
Query Match	73.7%; Score 2262.6; DB 6; Length 3084;			

Best Local Similarity 84.7%; Pred. No. 0;
Matches 2640; Conservative 0; Mismatches 399; Indels 78; Gaps 6;

```
QY 1 ATGCGCGCGCGCGTCAACGCGGACGACGAGCAACGAAATGACATTTGAGGAACAT 60
Db 1 ATGCGCGCGCGCGTCAACGCGGACGACGAGCAACGAAATGACATTTGAGGAACAT 60
QY 61 GTTTTACCTTTAATTCTAAAGAGATGSACTAAGTACGAGAAATGCAAAAAAATCTA 120
Db 61 GTTTTACCTTTAATTCTAAAGAGATGSACTAAGTACGAGAAATGCAAAAAAATCTA 120
QY 121 AAAAAATATTTGCCAGAAATGACTGAAGCAAACTAAATATAGAACAAATGACAGAAA 180
Db 121 AAAAAATATTTGCCAGAAATGACTGAAGCAAACTAAATATAGAACAAATGACAGAAA 180
QY 181 CTTAAAGCTTTTGGCGAAGATGGAAGAAACAGATACAAAATGCAAGAAATGCAAGCCAT 240
Db 181 CTTAAAGCTTTTGGCGAAGATGGAAGAAACAGATACAAAATGCAAGAAATGCAAGCCAT 240
QY 241 ATTGAGAAAAATGTACTACATCAAAAGSAAAATTAAAGAAATTTAAAAAAAATTT 300
Db 241 ATTGAGAAAAATGTACTACATCAAAAGSAAAATTAAAGAAATTTAAAAAAAATTT 300
QY 301 CAGATTATTAACGATTAAGATTTGCAAAAGAGATGAACAACATGCTATTTTGGAGGA 360
Db 301 CAGATTATTAACGATTAAGATTTGCAAAAGAGATGAACAACATGCTATTTTGGAGGA 360
QY 361 GTATGTTCAAAAAGAACTTTAAAGATGTTGCAATCTTTGAGAAATTAAGTGTATCAAAAG 420
Db 361 GTATGTTCAAAAAGAACTTTAAAGATGTTGCAATCTTTGAGAAATTAAGTGTATCAAAAG 420
QY 421 AAGCGTAAAGTGGCGAGAAAGTCTTTTAAGAGACTTGGTAGCATTTAAATGGA 480
Db 421 AAGCGTAAAGTGGCGAGAAAGTCTTTTAAGAGACTTGGTAGCATTTAAATGGA 480
QY 481 TCACTCATATGTGAAAAAAAATTAAAGATTTGCCCTGTCAATGGGAGGGAAGTAT 540
Db 481 TCACTCATATGTGAAAAAAAATTAAAGATTTGCCCTGTCAATGGGAGGGAAGTAT 540
QY 541 GAGTTTAACTTGTGTCTGACACGAGAAAGACATGTAAGATTTTAAATGAAAA 600
Db 541 GAGTTTAACTTGTGTCTGACACGAGAAAGACATGTAAGATTTTAAATGAAAA 600
QY 601 GATTAAGAGTGGCGGTACTTTAAACAGATGTTTACAGAGCACTAGGAAGTTTAAA 657
Db 601 GATTAAGAGTGGCGGTACTTTAAACAGATGTTTACAGAGCACTAGGAAGTTTAAA 657
QY 658 AAGAAAAATGTTGATCTGCAACGAGAAAGATGTAATTTTACATTTGAAATTTGCGAG 715
Db 658 AAGAAAAATGTTGATCTGCAACGAGAAAGATGTAATTTTACATTTGAAATTTGCGAG 715
QY 716 -----ACGACATATTAATTAATGT 735
Db 716 -----ACGACATATTAATTAATGT 735
QY 721 GACAAATCAAAAGTCAATTAACCTAATTAATTAAGACTGCAAAAGATATGTAACAGAGTGT 780
Db 721 GACAAATCAAAAGTCAATTAACCTAATTAATTAAGACTGCAAAAGATATGTAACAGAGTGT 780
QY 736 ATTAATTTGGAGGAAAAATGCCAAGAACAAATTTGCTTATATGCGCACGAGCCGAT 795
Db 736 ATTAATTTGGAGGAAAAATGCCAAGAACAAATTTGCTTATATGCGCACGAGCCGAT 795
QY 781 GATTAATTTAGCAAAAGTGTGAAAGAAATATGTTATATAGCATCCAGATCCGAT 840
Db 781 GATTAATTTAGCAAAAGTGTGAAAGAAATATGTTATATAGCATCCAGATCCGAT 840
QY 796 TTTGATCAACTAGAGCCGACGAGCTACATAGCAGAGATATAGGCTGGAAGATTTTAT 855
Db 796 TTTGATCAACTAGAGCCGACGAGCTACATAGCAGAGATATAGGCTGGAAGATTTTAT 855
QY 841 TTGCATCAACTAAGCCAGAGCTACACTAGCAGAGACATAGGCTGGAAGATTTTAT 900
Db 841 TTGCATCAACTAAGCCAGAGCTACACTAGCAGAGACATAGGCTGGAAGATTTTAT 900
QY 856 AAGAAAGTGAAGAGATGAGTTTATTTGAAAGAAATCATCTTAAGAGATGCAAGCT 915
Db 856 AAGAAAGTGAAGAGATGAGTTTATTTGAAAGAAATCATCTTAAGAGATGCAAGCT 915
QY 901 AAGAGGCGAAGAGATGAGTATTTTGTGGAAGACAAACATGTAAGAGATGCAAGCT 960
Db 901 AAGAGGCGAAGAGATGAGTATTTTGTGGAAGACAAACATGTAAGAGATGCAAGCT 960
QY 916 TTGTTGAGTGTGATGCAAGATTTCTAGTCTTAAAAAAGCAAGCAAGAGAAATGC 975
Db 916 TTGTTGAGTGTGATGCAAGATTTCTAGTCTTAAAAAAGCAAGCAAGAGAAATGC 975
QY 961 TTGTTGAGTGTGATGCAAGATTTCTAGTCTTAAAAAAGCAAGCAAGAGAAATGC 1005
Db 961 TTGTTGAGTGTGATGCAAGATTTCTAGTCTTAAAAAAGCAAGCAAGAGAAATGC 1005
QY 976 GAAAGAGCCCTTCAAAAAGTGCAGAAATCTCATGAAACATGAGGCTTTAAGAAATTTA 1035
Db 976 GAAAGAGCCCTTCAAAAAGTGCAGAAATCTCATGAAACATGAGGCTTTAAGAAATTTA 1035
```

```
Db 1006 ATAAAAACCCCTTAAAAAAAATCAGCAAAACCTCATGAAACATGAGCCTTAAAAATCTA 1065
QY 1036 TGTAAAGAAAAATGTTTAAATATGATGAACGAAAAATGTAAGAAATTTCAAAATGAT 1095
Db 1066 TGTAAAGAAAAATGTTTAAATATGATGAACGAAAAATGTAAGAAATTTCAAAATGAT 1125
QY 1096 ATTAACAAAATTTGCAAAATTTTCACTTCAAAAGTCACTAATAATGCTTTTGTGATCA 1155
Db 1126 GTTAACAAAATTTGCAAAATTTTCACTTCAAAAGTCACTAATAATGCTTTTGTGATCA 1185
QY 1156 ACAAAGAGAAATATGAAATTTGTTGATGGAAGAGGTTGGCAACATTTCTAGCAAGAA 1215
Db 1186 CC-----TATGAAATTTGGCAATTTGGGAAAAATTTACCAATTTCTTAGATGAA 1236
QY 1216 GATTGTCGAAATTTGGAGTCTATTTGTTCTATTTTGAATAAAAAATGTCAGATGAGAA 1275
Db 1237 GATTGTCGAAATTTGGAGTCTATTTGTTCTATTTTGAATAAAAAATGTCAGATGAGAA 1296
QY 1276 AATGATGTAAAAATATTAAGCAACATGTTACAAAAGAGAACTTGATGACGCGCAAT 1335
Db 1297 GAAAGCTGTATGATGATGAGGAGCGGTGTTACAGAGAGAGGCTTGATGACGCGCAAC 1356
QY 1336 AAGTGTGCAAGAAATATGCGAGAAATGTTACATGTTTCAAAACAAAGCTGCTTGA 1395
Db 1357 AGTGTGTTGCAAAAATATGCGGTGTTATGCTGTGTTCAATCAAACTTTGGCTTAAG 1416
QY 1396 AAGTTTCAACAAGAAATATGTAAGTAATGTAAGAACTGAAAAAAGAAACAAAGAAAT 1455
Db 1417 GAGTTTCAACAAGAAATATGTAAGTAATGTAAGAACTGAAAAAAGAAACAAAGAAAT 1473
QY 1456 TTCTCAACGATGAAATTAATTTATCTGTGTGTAACAGCAGCAAAAGCAGCGGTGCTT 1515
Db 1474 TTCCCAACGATGAAATTAATTTATCTGTGTGTAACAGCAGCAAAAGCAGCTTACTT 1533
QY 1516 ACACATATCTTGAATGAAACATATTTTACGACAAACATGATCAAAAGCGAGAT 1575
Db 1534 ACACATATCTTGAATGAAACATATTTTACGACAAACATGATCAAAAGCGAGAT 1593
QY 1576 TTCCCGACAGTAAAAATTTGCAAGAAATGAGGAGAAAGTGGCCAAAGTTTAAGAGAGAT 1635
Db 1594 TTTCGACAGATTAAGCTGCAAGAACTAGGAGAAATGCAAGATTTTAAGAGAGAT 1653
QY 1636 TCAAAAGAAATTAATCATGCCATGTCATACACTGAGCAGCAATGCAATGCGTTGGAGCT 1695
Db 1654 TCAAAAGAAATTAATCATGCCATGTCATACACTGAGCAGCAATGCAATGCGTTGGAGCT 1713
QY 1696 ACAGAAATTTTAAAGCAGTTTTTATGAGTGAACAAAGATTTTGAAGAACAAAGAA 1755
Db 1714 ACAGAAATTTTAAAGCAGTTTTTATGAGTGAACAAAGATTTTGAAGAACAAAGAA 1773
QY 1756 AGTTGTGTAAAAATACCTTAAAGAAAGTGTAAATGATGTTGAAGAGGAGATGACCGT 1815
Db 1774 AGTTGTGTAAAAATACCTTAAAGAAAGTGTAAATGATGTTGAAGAGGAGATGACCGT 1833
QY 1816 TTCTCTTTTATGTTGCTTCCAAAACGTTACGTTGAGCTGATGTTAAAGACGTGAAA 1875
Db 1834 TTCTCTTTTATGTTGTTTCCAAAACGTTACGTTGAGCTGATGTTAAAGACGTGAAA 1893
QY 1876 GACAGGTGTGAAGTATCAAAAAAATATTAAGCTTCATATATTAATTAATTTCTTGA 1935
Db 1894 GACAGGTGTGAAGTATCAAAAAAATATTAAGCTTCATATATTAATTAATTTCTTGA 1953
QY 1936 AATTAATCAATTAATAATAACAACCTGGAAGAAATTTGCTTGTGGCAATGATATGCG 1995
Db 1954 AATTAATCAATTAATAATAACAACCTGGAAGAAATTTGCTTGTGGCAATGATATGCG 2013
QY 1996 AATGATTTTCACTAATTTTCCAGTCTTACGAAAGATATGTTGATCAAAATCAAG 2055
Db 2014 AATGATTTTCACTAATTTTCCAGTCTTACGAAAGATATGTTGATCAAAATCAAG 2073
QY 2056 AAGCATGTGAGCCGTTCTATTAAGAAAGCCTTGAAGATGCTCTCAAAATGAGACTT 2115
Db 2074 AAGCATGTGAGCCGTTCTATTAAGAAAGCCTTGAAGATGCTCTCAAAATGAGACTT 2133
```


DG 1 TGCGCGGGCGGCTCAGCGGCGAGCAAAAGTGCACAGAAATAGCATTTATGAGAGCATG 60
QY 62 TTTTAGCTTTAATCTTAAAGGAGATGAGACTAAGTAGAGAGAGATGCAAAAATACTAA 121
DB 61 TTTTAGCTTTGATTTTAAAAAAAATGATTTGAAGTATCAAAAATGCAAACTTAAGTTGG 120
QY 122 AAAAATATTCGCAAGAAATGACTGAGCAAACTAAATATAGAACAAATACACAGAAAC 181
DB 121 AAGAAATTTGCAAAACATTAACAATGACAGATTTAAATCCAGAAAAGTTCCAGAAAAT 180
QY 182 TTTAAGCTTTTGGCAAGATGAAAAAGCATACAAATGCAAAAGAACTGAAAGCCATA 241
DB 181 TAAAGATTTTCTGATGATACGCGGAAACGAAATGAAAAATGTCAAGATCTAAAAAAG 240
QY 242 TTGAGAAAAATGACTACATCAAGGAAACCTTAAGAAAGCAATTAATAAAAAAATTC 301
DB 241 TCAATCAAAAATGATTAATTTCAAGGAAACCTTCAACGCTGCTAGAAAAAAAATTT 300
QY 302 AGATTTATACGAGATGAGATTTGCAAAAGATGAAACAATGCTATTTTGGAGGAG 361
DB 301 CAGAAATTAACAGATGAGAGATTTGCAAAAAGATGAAACAATGCTATTTTGGAGGAG 360
QY 362 TATGTTCAAAAGAACTTAAGATGATGCAATCTTTGAGAAATAGTCTATCAAAAGA 421
DB 361 CATGTCACAGAGAACTTAAGATGATGCAATTAATTAAGGAATACGTTATCAAAAAG 420
QY 422 AAGTGATTAAGTTGCGGAAAGATTTCTTTAAGAGACTTCGTAAGGATCTTAATGAT 481
DB 421 AACGGAACAATGTCGAGAAAGATTTCTTTGAGGCGCTTCGTAAGTCTCATATAA 480
QY 482 CAGTCATATGTAAGAAAAAATTTAAGAGATTTGCCCTGTCATGCGGAGGAAAGTATG 541
DB 481 CAAAGACATGTAAGAAAAAATGTAAGAGATTTGCCGGAATTTGAAGAAAGAAAGCATG 540
QY 542 AGTTAACAACTGTGTCTGAAACCAAGAAAGACATGTAAGATATTTTAATGAAGAA 601
DB 541 AATTAACGAGACTTGTCTTATCAAAAACAACATGCTTAAGTCTTGAACAAAGAA 600
QY 602 ATTAAGAGTCGCGGACTTAATAACAGATTTTCAAGAGCACTAGAAAGTTTAA--A 658
DB 601 AAGTAATATGATGATCTTGAAGAAAGATTAAGAAAGCACTTAAGAAAGATTAATGC 660
QY 659 AAGAAACATGCTCTGAATTTACTCGAACATGCTATTTTACATTTGAAATTTGCCGAG-- 715
DB 661 GAGAAAAATGCTACTATTTACTTTGAGCAATGTTACTTTCACAGAGGAACTGTGAAGAG 720
QY 716 -----ACGACGATTAATTAATGTA 736
DB 721 ACAAAATCAAGTGCAATTAACCTAATAATAAGACTGCAAAAGATTAATGTAAGTGTG 780
QY 737 TTTGAATTTGGAGAGAAATGCCAGAAACAATAATTTGCTTATATGCCCACAGAGACCGAT 796
DB 781 ATGAATTAAGCAAGAAAGTGTGAAAGAAATAATTTTATATGATCTACAGAGATCCGAT 840
QY 797 TTGATCCAACTAGGCGCAGAGGCTACAAATAGCAGAGATTAATGAGGCTGAAAGATTTTATA 856
DB 841 TCGATCCAACTAGGCGCAGAGGCTACCTAGCAGAGGACATGAGGCTGAAAGACCTTTATA 900
QY 857 AGAAGTGAAGAGAGATGAGATTTTATTTGAAAGAAATCATCTAAGAGATGCGACAGCT 916
DB 901 AGAGGCGAGAAAGAGATGAGATTTTGTGGAAGACAACATGTAAAGATGCAACAGCTT 960
QY 917 TGTGGCATTTCTGATCCAGATTTCTAGTCTTAATAAAAAAAGACAAAGAAAGAAATGCG 976
DB 961 TGTGGCATTTCT-----TCTTAAGAAACCTTTAAAAAGAAAGAAATGTA 1005
QY 977 AAGAAAGCTTTCAAAAAGTGCAGAAAATCTCTAGAACATGAGGCTTTTGAAGATTTAT 1036
DB 1006 TAAAGGCTTTTAAAAAAAATCTGCAAAACCTCATGAACTAGAGCTTTGAAATCTAT 1065
QY 1037 GTAAGAAAAATGTTTAATGTAATGAGAACGAAAGAAATGTGAAGAAATTTGCAAAATGATA 1096
DB 1066 GTAAGAAAAATTAACCAAGTGTGATGGAACGAAAAATGTGATGAAGTAAAAAAGATG 1125

QY 1097 TTAACAAAACCTTGCAAAATTTTCACTTCAAAAGTCACTAATAATGCTTTTGTGATCAA 1156
DB 1126 TTAACAAAACCTTGATCAAGCTTATCACTCAACAAATCTTAAAAACCGCTTTTACATTTCA 1185
QY 1157 CAAAAGAAATTAATGAATTTGTTGATGGAAGGTTGCCAATTTTCTTAACACAGAG 1216
DB 1186 C-----TGATGGAATTTGCCGAATGCGGAAATTAACCGACATTTCTTAATGATGAG 1236
QY 1217 ATTGTGGAATTTGAGTCTTATTTGTTTCTAATTTTGAAGAAAAATGTCAGATGGAAGA 1276
DB 1237 ATTGTGCAAACTGAAATCTTAATTTGCTTTATTAATAAGAACTTTGTCAGATGTAAG 1296
QY 1277 ATGATGTAATAAATATTAAGACAACTGTTACAAAGAGGACTTATGACAGGCAATA 1336
DB 1297 AAGCTTGTATGATGATGAGGCGAGCGTGTACAAAGAGGCGTTGATGACAGGCAACA 1356
QY 1337 AAGTGTGCAAGAAATATGTCGAGAAATGTTACATGTTCAACAAAGCTGCTTGA 1396
DB 1357 GTGTGTCGCAAAAATATATGCTGAGTTATTAACGTGTTCAAAAGTTGGCTTAAG 1416
QY 1397 AGTTTCAACAAGAAATTAAGTAAAGTATGAGAAACGMAAAAGAAACAAAGGAAGTT 1456
DB 1417 AGTTTCAACAAGATTAAGTAAAGTATGTAAGAGCT--AAAAGAAATTAAGGAAGTT 1473
QY 1457 TCTCAACGATGATTAATTTATTTCTGTGTGATACAGCCAGCAAAAGACCGGTTGCTTA 1516
DB 1474 TCCCAACGATGAAATATTTGTTCTGTGTGATACAGCCAGCAAAAGCTGCAAGATTTACTTA 1533
QY 1517 CACATGATCTTGAATGAAACATATCTTTTACGACAACAATGATCAAAAGCAGAT 1576
DB 1534 CACAGATCATCAATGAGGGTTACCTTTTACGACAACAATGATCAAAAGCAGAT 1593
QY 1577 TCCGCAAGATTAATAATTTGCAAGAAATTTGGGAGAAAGTCCAAAGATTTGAGAGAGAT 1636
DB 1594 TCCGCAAGATTAAGATGCAAGAAAGTCAAGGAAAGAAATCCCAAGATTTGGAAGAGAT 1653
QY 1637 CAAAAGAAATTAACATGCGCATGTCATACCTGAGCAGCATGCAATCGCTGGGGACTA 1696
DB 1654 CAAAAGAAATTAACATGCGCATGTCATACCTGAGCAGCATGCAATCGCTGGGGACTA 1713
QY 1697 CAGAAATTTTAAAGCAGGTTTATTTGATGATGAACAACAAGATCTTTGAAAGCCAGAA 1756
DB 1714 CAGAAATTTTAAAGCAGGTTTATTTGATGATGAACAACAAGATCTTTGAAAGCCAGAA 1773
QY 1757 GTTGTGTAATAATCTTAAGAAAGAAAGTGTAAATGCTCTAAGAAAGGAGATTAACCGTT 1816
DB 1774 GTTGTGTAATAATCTTAAGAAAGAAAGTGTAAATGCTCTAAGAAAGGAGATTAACCGTT 1833
QY 1817 TCTCTTTGTATGTCCTTCAAAAGCTACGTCGTCAGTCGATGCTAATAAGACGTGAAG 1876
DB 1834 TCTCTTTGTATGTCCTTCAAAAGCTACGTCGTCGTCGATGCTAATAAGACGTGAAG 1893
QY 1877 ACAAGTGTGAAGTATTAACAAAAATTAAGGCTTATTAATTAATTAATTTCTGA 1936
DB 1894 ACAAGTGTGAAGTATTAACAAAAATTAAGGCTTATTAATTAATTAATTTCTGA 1953
QY 1937 ATTAATTAATAAATTAACAACACTGGAAGAAATTTCCCTCTTGGCATACGTAATGCA 1996
DB 1954 ATTAATTAATAAATTAACAACACTGGAAGAAATTTCCCTCTTGGCATACGTAATGCA 2013
QY 1997 ATTAATTTACCTTAATTTGTCAGAGTCTTACGAAGAGAAATGTTGTAACAAATCAAGA 2056
DB 2014 ATTAATTTACCTTAATTTGTCAGAGCTTACGAAGAGAAATGTTGTAACAAATCAAGA 2073
QY 2057 AGCATGTCGAGCGGTTCTATTAAGAAAGGCTTTGAAAGTCTCTCAAGATGAGCTTC 2116
DB 2074 AGCATGTCGAGCGGTTCTATTAAGAAAGGCTTTGAAAGTCTCTCAAGATGAGCTTC 2133
QY 2117 AAGAAATTTGATGATTAATCTTAATGTAAGTGAACCTGATTTGAAAGATTTGTAAGTAG 2176
DB 2134 AAGAAATTTGATGATTAATCTTAATGTAAGTGAACCTGATTTGAAAGATTTGTAAGTAG 2193

QY 2177 CGGGAAGCTAATATATCGTCAATCAATGCGCTTATGCAAGCTTAACCAAGATACT 2236
DB 2194 CGGGAAAGGTAAATATGCTCAATCAATGCGCTTATGCAAGCTTAACCAAGATACT 2253
QY 2237 CTGGAAGAGTATGATGAGATGCTGTAAGAAAGAACTCTGTGAGAAATTTAGTGAAGAGTGG 2296
DB 2254 CTGGAAGAGTATGATGAGATGCTGTAAGAAAGAACTCTGTGAGAAATTTAGTGAAGAGTGG 2313
QY 2297 AAGAACAGTGCAGAAAGCATTTACCAACAGAAATTAAGCAACCGGACGCTGATCTAAAAAAG 2356
DB 2314 AAGAACAGTGCAGAAAGCATTTACCAACAGAAATTAAGCAACCGGACGCTGATCTAAAAAAG 2373
QY 2357 ATTATTAACATATTAAGAACTTAAGAAACGTGCGAGAGAAAGCAATGAACAGTCCAGTC 2416
DB 2374 ATTATTAACATATTAAGAACTTAAGAAACGTGCGAGAGAAAGCAATGAACAGTCCAGTC 2433
QY 2417 TTGTTTGTCACTCATTAAGAAAAAGAAAGTATGATCAAAAAAGTAATGCAAAAAACA 2476
DB 2434 TTGTTTGTCACTCATTAAGAAAAAGAAAGTATGATCAAAAAAGTAATGCAAAAAACA 2493
QY 2477 AGGATTAAGATGCGCTTTCAACCGACTTCAAGATACCAAAAAACATGTGAAAAATCTAC 2536
DB 2494 AGGATTAAGATGCGCTTTCAACCGACTTCAAGATACCAAAAAACATGTGAAAAATCTAC 2553
QY 2537 GAGGGGAGTTAAGATGTAATCCGTACAGAACTTAAGAACTTAAGCAATTTGATTTGGCAG 2596
DB 2554 GAGGGGAGTTAAGATGTAATCCGTACAGAACTTAAGAACTTAAGCAATTTGATTTGGCAG 2613
QY 2597 CAGAAATTTGAAGATATGATGATTTGAAGAAAGATGTAATTAATTTGATGATGAT 2656
DB 2614 CAGAAATTTGAAGATATGATGATTTGAAGAAAGATGTAATTAATTTGATGATGAT 2673
QY 2657 GCAGAAATTTGAAGATGATGATTTGAAGAAAGATGATGCAAAAAAGTAAATTAAGGCTT 2716
DB 2674 GCAGAAATTTGAAGATGATGATTTGAAGAAAGATGATGCAAAAAAGTAAATTAAGGCTT 2733
QY 2717 GTCCGAACTGTAAGAGCTCTGAGAGGTGAAGCCGCAAGAAACAGTGCAGAAAGTACAAAGA 2776
DB 2734 GTCCGAACTGTAAGAGCTCTGAGAGGTGAAGCCGCAAGAAACAGTGCAGAAAGTACAAAGA 2793
QY 2777 CAATCAACAAACAACAACCGTTGCCGATCCGAGGCAACGGAATGCAAAATCCTTAC 2836
DB 2794 CAATCAACAAACAACAACCGTTGCCGATCCGAGGCAACGGAATGCAAAATCCTTAC 2853
QY 2837 AGACAACAGACATGAGGTTTACACAGACATGCAACACAGCAGTCTACTATCAAT 2896
DB 2854 AGACAACAGACATGAGGTTTACACAGACATGCAACACAGCAGTCTACTATCAAT 2913
QY 2897 CTACCATCACTCAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2956
DB 2914 CTACCATCACTCAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2973
QY 2957 CGACAGAGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3016
DB 2974 CGACAGAGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3027
QY 3017 GGTGGAATGTGATGAGGGGGGTGATGATGATGATGATGATGATGATGATGATGATGAT 3072
DB 3028 GGTGGAATGTGATGAGGGGGGTGATGATGATGATGATGATGATGATGATGATGATGAT 3083

RESULT 7
AR438436 3090 bp DNA linear PAT 20-FEB-2004
LOCUS AR438436
DEFINITION Sequence 5 from patent US 6664053.
ACCESSION AR438436
VERSION AR438436.1 GI:42663291
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3090)
AUTHORS Kovacs,J.A., Huang,S., Masur,H., Fischer,S.H., Gill,V.J. and Mei,Q.

TITLE Identification of a region of the major surface glycoprotein (MSG)
JOURNAL gene of human Pneumocystis carinii
FEATURES Patent: US 6664053-A 5 16-DEC-2003;
SOURCE Location/Qualifiers
1. 3090
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 61.9%; Score 1900.6; DB 6; Length 3090;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 609; Indels 84; Gaps 8;

QY 1 ATGCGCGGGCGGTCAAGCCGCGAGGCGAGGAGACAGAAATAGCATTTGATGAGAACT 60
DB 1 ATGCGCGGGCGGTCAAGCCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 60
QY 61 GTTTAGCTTTAATTTCTTAAGAGATGAGCTAAGTGAAGAGATGAGAGATGAGAGATGAGAGAT 120
DB 61 GTTTAGCTTTAATTTCTTAAGAGATGAGCTAAGTGAAGAGATGAGAGATGAGAGATGAGAGAT 120
QY 121 AAAAAATATGCGCAAGATGAGTGAAGCAAACTAATATATGAACAAGTACAGAGAAA 180
DB 121 AAAAAATATGCGCAAGATGAGTGAAGCAAACTAATATATGAACAAGTACAGAGAAA 180
QY 121 GAAGATATGCGCAAACTTAACAATGCAAGATTAATCCAGAAAAAGTTACAGAAAA 180
DB 181 CTTAAGGTTTTCGGAAGATGAGAAAGCAGATACAAATATGCAAAAGACTGAAGCCAAAT 240
DB 181 TTAAGAGATTTCTGTGTAACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 241 ATTGAGAAAAATGATCAATCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 241 GTCAATCAAAATATCAATTAATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 301 CAGATTAACGATTAAGATGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 301 TCAGATTAACGATTAAGATGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 361 GTATGTTCAAAAGACTTAAGATGATGCAATCTTTGAGAAATTAAGTCTATCAAAAG 420
DB 361 GCATGTCACACAGAACTTAAGATGATGCAATCTTTGAGAAATTAAGTCTATCAAAAG 420
QY 421 AAAGGTGTAAGTGGGAG 480
DB 421 GAACGAGAACTATGAG 480
QY 481 TCAGTCAATGTGAAG 540
DB 481 ACAAGAGCATGTGAAG 540
QY 541 GAGTTAACAACTTGTGTCTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 541 GAATTAACGAGCTTTGTCTTATCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 601 GATAAGAGTGCAG 657
DB 601 AAGAGTAATGTGATCTCTTGAAG 660
QY 658 AAG 715
DB 661 CGAG 720
QY 716 -----ACGACGATATTAATTAATGT 735
DB 721 GACAAATCAAGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGT 780
QY 736 ATTGAATGGAG 795
DB 781 GATGATTAAG 840
QY 796 TTGATCCAACTAAGGCGCAG 855
DB 841 TTGATCCAACTAAGGCGCAG 900

QY 856 AAGAGGTAGAGAGATGAGTTTATTTGAGAAAGATCATCTAAGAGATGCAGACT 915
 DB 901 AAGAGGCGAGAGAGCATGAAATTTTGTGAGAGACAAACATGTAAGAGATGCACGCT 960
 QY 916 TTGTTGGCATTTGTGATCCAAAGATTCTAGTCTTAAAAAAGACGACAAAGACAATGC 975
 DB 961 TTGTTGGCATCTACT-----TCTTAAGAAACCTTAAAAAAGAAAGATGT 1005
 QY 976 GAGAGAGCCCTTCAAAAAAGCTGCAAAAAATCCTCATGAAACATGAGGCTTTGAAAGTTTA 1035
 DB 1006 ATAAAGCCCTTAAAAAAGCTGCAAAAAATCCTCATGAAACATGAGGCTTTGAAAGTTTA 1065
 QY 1036 TGTAAAGAAAAATGCTTTAAGTAATGATGAGAAAGAAAAATGTAAGATTCGCAAAATGAT 1095
 DB 1066 TGTAAAGAAAAATTAACCAAGTAGATGATGAGAAAGAAAAATGTAATGAACTAGAAAAAGAT 1125
 QY 1096 ATTAACAAAAATCTGCAAAAATTTCACTTCAAAAAGTCACTAATAATTCGCTTTTGAATCCA 1155
 DB 1126 GTTAACAAAAATCTGTACAAAGTCTTAACATCAAAATCTTAAAAACCGCTTTTACATTTCA 1185
 QY 1156 ACAAAAGAAATTAATGAATTTGATGAGAGGAGGTTCCCAACATTTCTTAGCAAGAA 1215
 DB 1186 CC-----TGATGAAATTTGCGAATGGGAAAAATTAACGACATTTCTTAGATGAA 1236
 QY 1216 GATTGTCGAAATTTGAGTCCCTATTTGTTCTATTTTGAAAAAAATGTCAGATGAGAA 1275
 DB 1237 GATTGTCGAAAACTAGATCTTATTTGCTTTATTAAGAAACCTTGCCAGATGTCAAA 1296
 QY 1276 AATGCATGTAAAAATATTAAGACCAATGTTTACAAAAGAGACCTTGATGACGGGCAAT 1335
 DB 1297 GAAAGCTGTATGATGTGAGGGGACGCGTGTATTAAGAGGGCTTGATGACGGGCAAAAC 1356
 QY 1336 AAGTGTGCGAAGAAATATGCGAGGAATGTTACATGGTTCAAACAAAAGCTGGCTTGA 1395
 DB 1357 AGTGTGTTGCAAAAATATGCGAGGCTTATTCATGGCTCAAAATTAAGATGGCTTTAAG 1416
 QY 1396 AAGTTTCAACAGAAATTAAGTAAGTAATGTCAGAAACCTGAAAAAGAAACAAAGAAAT 1455
 DB 1417 AAATTTCAACAGAAATTAAGCAAAAGTAATGTGAAACG---AAAGAAATTAAGAAAT 1473
 QY 1456 TTCTCAAAACGATGAATTAATTTATTTCTGTGTACAGCCAGCAAAAGACCCGGTTGCTT 1515
 DB 1474 TTCTCGAAGCATGATTTGTTGTTCTGTGTATACAAACGACAAAGGCGACGATTTCTT 1533
 QY 1516 ACACATATCTTGAATGAAGAACTATCTTTTACGACAAACATGATCAAAACGAGAT 1575
 DB 1534 ACACATCACCATCAAAATGAGGTTATCTTTTACGACAAACATGATCAAAACGAGAT 1593
 QY 1576 TTCCCGACAGATAAAATTTGCAAGAAATTTGGGAGAAAGTCCCAAGATTTTAGAGAGAT 1635
 DB 1594 TTTCGCAAGATTAAGACTGCAAGAAATTTAGGAGAAATTCCAAGATTTTAGAAAGAT 1653
 QY 1636 TCAAAAGAAATTAATGAGCCATGTCATACCTGAGACAGCAATGCAATTCGCTTGGGACT 1695
 DB 1654 TCAAAAGAAATTAATGAGCCATGTCATACCTGAGACAGCAATGCAATTCGCTTGGGACT 1713
 QY 1696 ACAGAAATTTTAAAGCAGGTTTTTATGAGTGAACAACAAGATTAATCTTGAAGACCAAGAA 1755
 DB 1714 ACAGAAATTTTAAAGCAGATTTTATGAGTGAACAACAAGATTAATCTTGAAGACCAAGAA 1773
 QY 1756 AGTTGTATAAATCTTAAAGAAAGTGTATAATGCTTAGAAGAGAGATGACCGT 1815
 DB 1774 AACTGTCAAAAATTTTAAAGAAAGAAATGCAATAATGCTTAGAAGAGAGATGACCGT 1833
 QY 1816 TTCTCTTTTGTATGTCTTCAAAAACGCTACGTTGAGCTGATGTTAAAGACGTGAAA 1875
 DB 1834 TTTTCTTTTGTATGTCTTCAAAAACGCTACGTTGAGCTGATGTTAAAGACGTGAAA 1893
 QY 1876 GACAGGTGTGAAGTATCAAAAATTAATAAAGCTTCAATATTAATTTGAAATTTCTTGA 1935
 DB 1894 GATAGGTGTGAAGTATCAAAAATTAATAAAGCTTCAATATTAATTTGAAATTTCTTGA 1953
 QY 1936 AATATATCAAAATTAATAAATCAACCTGGAAGAAAGAAATTTGCTTGTGCAATACGTAATTC 1995

DB 1954 AAAAATCAAAATTAAGAGAGATCAGCAGCAAAATTTTGTCCCTCATGCACTCCATCTGC 2013
 QY 1996 AATGATTTTCACTAATTTGTCCAGGCTTACGAAAGAGAAATGT---TGTACAAAATTC 2052
 DB 2014 GATGATTTTATCCCAATTTGTCTGATCTTAAGAAAGAAATTTTCTGTCAAAATCTT 2073
 QY 2053 AAGAGCATGTGAGCCGTTCTATTAAGAAAGAGCCCTTGAAGATGCTTCAAGTAGAG 2112
 DB 2074 AAAAAATATTGCGAACCATCTACAAAAGAAAGTTTAGAAGATGCTTTAAAGTAGAG 2133
 QY 2113 CTTCAAGAAAATTTGACTGATTAATCTTAAATGTGAACCTTGATTTGAAAGATTTGTACA 2172
 DB 2134 CTTGAGAGAAATTTAATTAATATTAATTAATGTGAACCTGATTAAGAAAGATTTGTACA 2193
 QY 2173 GTAGGGGAAACGTAATTAATGCTCAATGAGGCTTATGCAAGGCTAACCAAGAGAT 2232
 DB 2194 GTATTGAAGCGTAATTAATGCTCAATGAGGCTTATGTAAGATTAATCAAGAAAT 2253
 QY 2233 AACTCTGAAAAG--AGTATGAGATGCTAAGAAAGAACTGTGTGAAATTAAGTAAA 2289
 DB 2254 AAAAATTAAGAGCCGATTAATTAAGATTTAGAAAGAAAGCTTTGTCTTAAATTAAGTGA 2313
 QY 2290 GAAGTGAAGAACGTGCAAAACATTTACAAACAGATTTAGCAACCGGCACTGATCTA 2349
 DB 2314 GAGGTGAAGAACGATCAAAAGTATTACCAACAGAAATTAACAGAGCTGGAAGAAAGCTTA 2373
 QY 2350 AAAAAATTAATTAAGATTAAGAGAACTTAAGAAAGTCAAGAGAAAGCAATGAACAAG 2409
 DB 2374 AAAAAATATTTAAGATTAAGAGAACTTAAGAAAGGCAAAAAAAGCAATGAACAAG 2433
 QY 2410 TCCAGCTTTGTTTGTCACTCATTAAGAAAAAGAAAGTAATGTATCAAAAATTAAGTAC 2459
 DB 2434 TCCAGCTTTGTTTATCACTTTGTTAAGAAAAAGAAAGTAATGTATCAAAAATTAAGTAC 2493
 QY 2470 AAAAAAGAGATTAAGATTAAGAGATTTGCAACGACTTCAAGATTAACCAAAACATGTGAA 2529
 DB 2494 AAAAAAGAGATTAAGATTAAGATTTGCAACGACTTCAAGATTAACCAAAACATGTGAA 2553
 QY 2530 ATACTAGAGAGGAGTAAAGATTAAGATTTGCAACGACTTCAAGATTAACCAAAACATGTGAA 2589
 DB 2554 ATACTAGAGAGGAGTAAAGATTTGCAACGACTTCAAGATTAACCAAAACATGTGAA 2613
 QY 2590 TTGGCAGCAGAGATTTTGAAGATTAAGATTTTGAAGAAAGATTAATTAATTTGAA 2649
 DB 2614 TTGGCAGCAGAGATTTTGAAGATTAAGATTTTGAAGAAAGATTAATTAATTTGAA 2673
 QY 2650 TCAAGATTTGCAAGATTAAGAGATTTGCAAGACTTGAAGAGATTAATTAATTTGAA 2709
 DB 2674 TCGGATTTGCGGATTAAGAGATTTGCAAGACTTGAAGAGATTAATTAATTTGAA 2733
 QY 2710 AAGCTTTGTGCAATTTGAAGCTTTGAGAGTGAAGCCGCAAGAAAGATTAATTTGAA 2769
 DB 2734 AAGCATGTGACGATTTGAAGCTTTGAGAGTGAAGCCGCAAGAAAGATTAATTTGAA 2793
 QY 2770 ACAACGACAACTACAAACAAACAAACGCTTCCGATCCGAAGGCAAGCAAGATTCGAA 2829
 DB 2794 ACAACGACAACTACAAACAAACAAACGCTTCCGATCCGAAGGCAAGCAAGATTCGAA 2853
 QY 2830 TCTTTACAGACAAACAGACATGAGGTTTACAGACATGACACACACAAAGCCTTACT 2889
 DB 2854 TCTTTACAGACAAACAGACATGAGGTTTACAGACATGACACACACAAAGCCTTACT 2913
 QY 2890 ATCATATCTACATCAATCAAAATTAACATTTGACATCAACGAGGCAATGCAACCAAC 2949
 DB 2914 ATCATATCTACATCAATCAAAATTAACATTTGACATCAACGAGGCAATGCAACCAAC 2973
 QY 2950 AAGGTACGACAGGAGAGAAATGATGACGAGAGACGTGAAACCGAGTGAAGGCTGAGG 3009
 DB 2974 AAGGTACGACAGG-----GATGAAGCAGAGACGTGAAACCGAGTGAAGGCTGAGG 3027
 QY 3010 ATGAGGTGTGAATGTGATGAGGGGGGTATGATGCAATGTTATTTGTTTCTTATGATTT 3069

Db	3028	ATGATGGGTGGAGCGTATATAGAGGGGGGTGATAGACGAAGTATTATTCGTTATGATT	3087
Qy	3070	TAG 3072 	
Db	3088	TAG 3090	
RESULT 8			
LOCUS	AF033208	3089 bp	DNA linear PLN 10-SEP-1998
DEFINITION	Pneumocystis carinii f. sp. hominis clone HUMSG11 major surface glycoprotein (MSC) gene, partial cds.		
ACCESSION	AF033208		
VERSION	AF033208.1	GI:3560512	
KEYWORDS			
SOURCE	ORGANISM		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBLISHED			</

	ORIGIN	
	Query Match	61.8%; Score 1899, 6; DB 8; Length 3089;
	Best Local Similarity	77.8%; Pred. No. 0;
	Matches 2429; Conservative	0; Mismatches 609; Indels 84; Gaps 8
Oy	2 TGGGCGGGGGGGTCAAGCGCAGGCGAGGACAACAGATATGACTTGATGAGGAACATG	61
Dd	1 TGGCGCGGGCGGTCAAGCGCGGGCAAAAGGTGCACAAATAAGCATTTATGAGGAGCATG	60
Oy	62 TTTTAGCTTTAATTCTAAAGGAATGGACTTAAGTAGCACAGAATGCAAAAAAACTPA	121
Dd	61 TTTTAGCTTTGATTAAAABAAAATGGATTGAAGATPACAAAATGCCAAAATTAA GTTG	120
Oy	122 AAAAATATTCGCAAGAAATTGACTGAGCAAAACTPAATATGAAACAATACACAGAAAC	181
Dd	121 AAGATATATTGCAAAAACCTTAACAAATGCAGGATTTAAATCCGAAAAAGTTTACGAAAAAT	180
Oy	182 TTAAAGGTTTTTGGGAAGATGGAAAACGAGTACAAAATGCAAAAGAACGTGAAGCCAAAT	241
Dd	181 TAAAAAGATTTCTGATATACGGGAAACGAAATGAAAAATGTCCAGATCTTAAAAAACAAAG	240
Oy	242 TTGAGAAAAAATGTACTACATCAAAAGAAAACTTAAAGAAGCAATTAABAAAAAATTC	301
Dd	241 TCATTCAAAAATGCATTAAATTTCAAGSAAACCTTCAACAGCTGCTGGAAAAAATTT	300
Oy	302 AGATTATPACGATPAAGGATTGCCAAAGAAATGAACAAATGCTATTTTGGAGGAG	361
Dd	301 CAGATTATPACAGATGAGGATTGCCAAAAAGAAATGAACAAATGCTATTTTGGAGGAG	360
Oy	362 TAATGTTCAAAAGAACTTAAAGATGATTCGCAATCTTTGAGAAATPAGTCATCAAAAGA	421
Dd	361 CATGTCCAAACGAACCTTAAAGATATCGCAATTAATTAAGAAATPACGTTATCAAAAAG	420
Oy	422 AACGTGATPAAAGTTGCGGAAGAAGTCTTTTAAAGGCACTTCGTAGCATCTTAATGAT	481
Dd	421 AACGGAACMATGTGGCAGABAAGTCTTTTGAAGGCGCTTCGTGTGATCTCAATGA	480
Oy	482 CAGTCATATGTGAAAAAAAACCTTAAAGAAATTTGCCCTGTATGCGGAGGGAAGTGATG	541
Dd	481 CAAGAATCATGTGAAAAAAAAGCTGAAGAAGTTGGCCCAAAATTTAGAAAGAGAAACGATG	540
Oy	542 AGTTTACAAACTTGTGTCTGAACCAAGAAAGACATGTAAAGATTTTAAATTTGA	601
Dd	541 AATTATPCGAGACTTGTCTTTATCAAAAAMCAACATGTGTAAAGTCTTTATCAAAAAGGA	600
Oy	602 ATTAAGAAGTGGGTACTCTTAAACAGATGTTTTGACGACACTAGGAAGTTTTAA---A	658
Dd	601 AAAGTAAATGTGATACCTTTGAAAAAGAAAGTTGAAGAACACTTAAAGAAATGAATTC	660
Oy	659 AAGAAACATGTCTTGAATTTACTCGAACATCTATTTTTCATTTGAAATTCGCGAG---	715
Dd	661 GAGAAAAATGTCTTATTTACTTTGAGCAATTTTACTTTTCAACAGAGGAACGTGAAGAG	720
Oy	716 -----ACGACGATTAATTAATGTA	736
Dd	721 ACAATACAAATGCAATTAACCTTAATATAAAGACTGCAGAAAGAAATGTACCAGAGTGTG	780
Oy	737 TTGAATGGGAGGAAAAATGCCAAGAACAAATATTGCTTATATGCGACACAGSACCCGAT	796
Dd	781 ATGAATTTAGCAAAAAAGTGGAAAAAGAAATATTGTTTATATGTATCCAGGATCCGATT	840
Oy	797 TTGATCCAACTAGGCCACAGAGCTTCAATATAGCAGAGATATATAGGCTGGAAGATTTTATA	856
Dd	841 TCGATCCAACTTAGCGCAGAGCTTCACTATAGCAGAGGACATATAGGCTGGAAGACTTTATA	900
Oy	857 AGAAGTATAGAGAGATGAGATTTTATTTGAAAAAATCATCTTAAGATGCGACAGCTT	916

Db 901 AGAGGGCAGAAAGAGATGGAATTTTGTGGAAAGCAATGATGAGATGCACAGCTT 960
Qy 917 TGTGGCATTTGTGATCCAGATTTAGTCTTTAAAAAAGAGACAAAGAAATGCG 976
Db 961 TGTGGCCTACT-----TCTTAAAGAAACCTTTAAAAAAGAAAGTGA 1005
Qy 977 AAGAGCCCTTCAAAAAAGTCGCAAAAAATCCTCATGAATGAGGCTTAGAAAGTTAT 1036
Db 1006 TAAAGCCCTTAAAAAAGTCGCAAAAAATCCTCATGAATGAGGCTTAGAAATCTAT 1065
Qy 1037 GTAAGAAAAATGCTTAAAGTAATGATGAGACGAAAAATGTGAAGATTTGCAAAATGATA 1096
Db 1066 GTAAGAAAAATTAACCAAGTATGATGAGACGAAAAATGTGAATGATACTAGAAAAAGATG 1125
Qy 1097 TTAACAAAACCTTGCAAAATTTTCACTTCAAAAAGTCACTAATATGCTTTTGTATCCA 1136
Db 1126 TTAACAAAACCTTGATCAAGTCTTACATCAACAAATTTTAAAAACGCTCTTATCAATTTCA 1185
Qy 1157 CAAGAAGAAATTAATGAATTTGTGAATGGGAAGGTTGCAACATTTCTTAGCAACGAG 1216
Db 1186 C-----TGAATGAATTTGCGAATGGGAAATTTCCGACATTTCTTAGTGAATGAG 1236
Qy 1217 ATTTGCGAAATTTGAGTCTTATTTCTATTTTGAAAAAAATGTCAGATGAGAA 1276
Db 1237 ATTTGCGAAATTTGAGTCTTATTTCTTATTTAAAGAAACCTTGCAGATGTCAGAG 1296
Qy 1277 ATGCAATGTAATTAATTAAGACACATTTACAAAAGAGACTTATGACCGGCAATA 1336
Db 1297 AAGCTTGTATTAATGTGAGGCGAGCGTGTATTAAGAGAGGCTTGATCACGGGCAACA 1356
Qy 1337 AAGTGTGCAAGAAATTAATGAGGAGATGTTACATGTTCAAAACAAAGCTGGCTGAA 1396
Db 1357 GTGTTGTGCAAAAAATTAATGAGGAGGTTATGTCAGGCTCAAAATTAATGAGCTTAAGA 1416
Qy 1397 AGTTTCAACAAGATTTAGTAAAGATGATGAGAAACCTGAAAAAAGAAAAACAAGAAAT 1456
Db 1417 AATTTCAACAAGATTTAGCAAAAGTATGAGAAATG---AAAGAAATTAAGAAAT 1473
Qy 1457 TCTCAACGATGAATTAATTTTCTGTGTGACACCCGCAAAAGAGCCGGTTGCTTA 1516
Db 1474 TCTGAAACGATGAATTTGTTCTGTGTATTAACAACGAAAGGACGACAGATTAATCTTA 1533
Qy 1517 CACATGATCTTGAATGAAGAACTATCTTTTACGACAACTGGATCAAAAGGAGAT 1576
Db 1534 CACATCACATCAATGAGATTAATCTTTTACGACAACTGGATCAAAAGGAGAT 1593
Qy 1577 TCCCGACAGATTAATAATGCAAGAAATGGGAGAAATGGCCAAAGATTAGAGAGAT 1636
Db 1594 TTCCGACAGATTAAGATGCAAGAAATGGGAGAAATGGCCAAAGATTAGAGAGAT 1653
Qy 1637 CAAGAAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1696
Db 1654 CAAGAAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1713
Qy 1697 CAGAAATTTTAAAGCAGGTTTATTTGATGAAACAAAGATATCTTGAAGAACCAAGAA 1756
Db 1714 CAGAAATTTTAAACAGATTTTATTTGATGAAACAAAGATATCTTGAAGAACCAAGAA 1773
Qy 1757 GTTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1816
Db 1774 ACTGTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1833
Qy 1817 TCTCTTTTGTATGTCTTCCAAAACGCTACGCTGATGATGATGATGATGATGATGATGAT 1876
Db 1834 TTTCTTTTGTATGTCTTCCAAAACGCTACATGATGATGATGATGATGATGATGATGAT 1893
Qy 1877 ACAGGTGAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1936
Db 1894 ATAGGTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1953
Qy 1937 ATAAATACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1996

Db 1954 AAAATCAATTAAGACAAATCAACAGCAATATTTGCTCATGATCCATACATGCG 2013
Qy 1997 ATAGATTTTACCCATATTTGTCAGGCTTACGAAGAAGAAATGAT---TGTACAAAAATCA 2053
Db 2014 ATAGATTTTACCCATATTTGTCAGGCTTACGAAGAAGAAATTTTCTGTCAAAATCTTA 2073
Qy 2054 AGAAGCATGCTGAGCCGCTTATTAAGAAAGAGCCTTGAAGATGCTCTCAAGATAGAGC 2113
Db 2074 AAAAATTTGGAACCATTTCTACAAAAGAAAGTTTGAAGATGCTCTTAAGATAGAGC 2133
Qy 2114 TTCAAGAAAAATGATGATTAATCTAATGTAACCTGCACTTGAAGAAAGATTTGTACAG 2173
Db 2134 TTCAAGAAAAATTTAAGTAATTAATCTAATGTAACCTGCACTTGAAGAAAGATTTGTACAG 2193
Qy 2174 TAGGGGAAACGTAATTAAGCCGCAATGATGCTTATGCAAGCTTAACCAAGGATA 2233
Db 2194 TATTGAAGAGTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2253
Qy 2234 ACTCTGAAAG---AGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2290
Db 2254 AAATCTAAAAAGCCGATTAATTAATTTGAAGAAAGCTTTGTCTAATTAATGAGAG 2313
Qy 2291 AAGTGAAGAACAGTGCAGAAAGCATTAACAGAAATTAAGACAAACCGGACGTAATCTAA 2350
Db 2314 AGTGAAGAACAGCAATGCAAAAGTATTAACAGAAATTAACAGACGTAAGAAAAAGTCTAA 2373
Qy 2351 AAAAAGATTAATTAAGCATTAATGAGAACTTAAAGAAACGTCGAGAGAGCAATGAACAAT 2410
Db 2374 AAAAAGATTTAAGCAATTAATGAGAACTTAAAGAAAGGCAAAAAAGCAATGAACAAT 2433
Qy 2411 CCACTTTGTTTGTCACTCACTTAAGAAAAAGAAATTAATGATCAAAAAAGTAAATGACA 2470
Db 2434 CCACTTTGTTTGTCACTTTGATTAAGAAAAAGAAATTAATGATCAAAAAAGTAAATGACA 2493
Qy 2471 AAAACAAAGATTAAGATGCGGTTTCAACGCACTTCAAGATTAACCAAAAAATGTAAGAA 2530
Db 2494 AAAACAAAGATTAAGATGCGGTTTCAACGCACTTCAAGATTAACCAAAAAATGTAAGAA 2553
Qy 2531 TACTTACGAGGAGGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2590
Db 2554 TACTTACGAGGAGGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2613
Qy 2591 TGGCAGCAGAAATTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2650
Db 2614 TGGCAGCAGAAATTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2673
Qy 2651 CAGATTCAGAAATTAAGAGGATGCAAAAGCTTGAAGAAAGTATGCAAAAAAGATTAATA 2710
Db 2674 CGATTTGCGGATTAAGAGGATGCAAAAGCTTGAAGAAAGTATGCAAAAAAGATTAATA 2733
Qy 2711 AGGCTTGTGCAATCTGAGGCTTGTGAAGTGAAGCCGACAGAAACGTGACAGAAAGTA 2770
Db 2734 AGATATGTCAGATCTGAAGCTTGTGAAGTGAAGTGTGCAATGAAATGTGTCACAGAAAGCA 2793
Qy 2771 CAACGAACTAAGCAACAAACAAACAAACGCTGCGAATCCGAAGCAACGGAATGCAAT 2830
Db 2794 CAACGAACTAAGCAACAAACAAACAAACGCTTACCAATCCGAAGCAACAGAAATGCAAT 2853
Qy 2831 CTTTACAGACAAACAGACATGAGGTTACAGACATGACACACACAGACAGCTCTACTA 2890
Db 2854 CTTTACAGACAAACAGATTAATGAGGTTACAGACATGACACACACAGACAGCTCTACTA 2913
Qy 2891 TCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2950
Db 2914 TCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2973
Qy 2951 AGTGTACGACGAGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3010
Db 2974 AGTGTACGACGAG-----GATGAAGCAGAGACGTAAGACGATGAGGATTTGAAGA 3027
Qy 3011 TGAATGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3070
Db 3028 TGAATGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3087

Db 1389 GCAACAGTACTAAATTTCTTCTGAAAAATAAGATAAATCTCTCT----- 1339
Oy 1159 AAAGAAATATATGAATTTGGATGGGAGGTTGCCACATTTCTTAGCAAGAAAT 1218
Db 1340 ---GGAATCCGAGAAACCATTCATGTATGAATGTGACATTTCTTAGCAAGTAC 1336
Oy 1219 TGTGCGAAATTTGAGTCCCTATTTGTTCTATTTTGAATAAAATATGTCAGATGAGAAAT 1278
Db 1397 TGTGCAAGTTAGATCAGACTGTTTTTATTTTGTCTCAAGATTAAGATCCCTTAATAAA 1456
Oy 1279 GCATGTAAAAATATAGAGCAATGTATCAAAAGAGACTTGTATGACGGGCAATTA 1338
Db 1457 GAATGCAAGAAATGTTAAGGAGCAATGTATTAAGAGGGCTTGAATGACGGGCAATTA 1516
Oy 1339 GTGTGCAAGAAATATGCGAGAAATGTATCATGTTCAACAAAGCTGGCTGAAG 1398
Db 1517 GTGTGCAAGAAATATGCGAGAAATGTATCATGTTCAACAAAGCTGGCTGAAG 1576
Oy 1399 TTTCAACAGAAATTTAGTAAAGTATGTAGAAACTGAAAAAGAAACAAAGAAATTT 1458
Db 1577 TTTCAACAGAAATTTAGTAAAGTATGTAGAAACTGAAAAAGAAATTAAGAAATTT 1636
Oy 1459 TCAACAGATGAATTTATTTATTTCTGTGTGTACAGCCAGCAAAAGACCCGGTGTCTA 1518
Db 1637 TCAACAGATGAATTTATTTGTTCTGTGTGTACAGCCAGCAAAAGACCCGGTGTCTA 1696
Oy 1519 CATATCTTGAATGAATACTATCTTTTATAGCAACAACTGATTCAAAGCGATTC 1578
Db 1697 CATATCTTGAATGAATACTATCTTTTATAGCAACAACTGATTCAAAGCGATTT 1756
Oy 1579 CCGACAGATTAATAATTTGAGAAATTTGGGAGAAATTTGCCAAGATTTTGAAGAGAT - 1636
Db 1757 CCGACAGATTAATAATTTGAGAAATTTGGGAGAAATTTGCCAAGATTTTGAAGAGATTC 1816
Oy 1637 -CAAAAGAAATTAATGAGCAATGTCTATCACTGAGCAGCAATGCAATCGCTTGGGACT 1695
Db 1817 CAAAAGAAATTAATGAGCAATGTCTATCACTGAGCAGCAATGCAATCGCTTGGGACT 1876
Oy 1696 ACAGAAATTTTAAAGCAGTTTATTTGATGAACAAAGATTAATCTTTGAAAGCAAGAA 1755
Db 1877 ACAGAAATTTTAAAGCAGTTTATTTGATGAACAAAGATTAATCTTTGAAAGCAAGAA 1936
Oy 1756 AGTGTGTAATAATCTTAAAGAAAGTGTATTAATGCTTAGAAGAGAGATGACCGT 1815
Db 1937 AACGTGTAAATTTTAAAGAAAGTGTATTAATGCTTAGAAGAGAGATGACCGT 1996
Oy 1816 TTCTCTTTTGTATGTCTTCAAAACGCTACGTGTGAGTGTATGTAAGAGAGCTGAA 1875
Db 1997 TTCTCTTTTGTATGTCTTCAAAACGCTACGTGTGAGTGTATGTAAGAGAGCTGAA 2056
Oy 1876 GACAGGTGTGAATTTTAAAGAAATTTAAAGCTTATATTTATTTGAAATTTCTTGA 1935
Db 2057 GATGAGTGTGAATTTTAAAGAAATTTAAAGCTTATATTAAGAAATTTGTTCTTAA 2116
Oy 1936 AATATATCAATTAATAATCAACTGGAAGAAATTTGCCCTTGGCATACGATTC 1995
Db 2117 AATATATCAATTAATAATCAACTGGAAGAAATTTGCCCTTGGCATACGATTC 2176
Oy 1996 AATAGATTTTACCTAATTTGTCCAGTCTTACGAAAGAAATAGTTGTACAAATCAAG 2055
Db 2177 AATAGATTTTACCTAATTTGTCCAGATTTTCAAAAAGAAATCTTGTACAAATTA 2236
Oy 2056 AAGCATGCTGAGCGCTTCTATTAAGAAAGCTTGGAAAGTGTCTCTCAAGTGAAGTT 2115
Db 2237 AACATTTGTAAAGCAATTTTAAAGAAAGCTTGGAAAGCTCTCAAGTGAAGTT 2296
Oy 2116 CAAGGAATTTGATGATTAATCTTAATGTGAACCTGATGAAAGATTTGTACAGTA 2175
Db 2297 CGAGGAAGCTGAGTGAAGTGAAGAAATCTACTGACATTTGAGGAGATTTGTACACTA 2356
Oy 2176 GCGGGAACGTAAATATGCGTCAATCAAGTGGCTTATGCAAGCAAGCAAGATTAAC 2235
Db 2357 GCGGGAATTTAAATATGAGCATCAAGTGAAGTATTAATCAAGATTAACATCAAGTATGC 2416

Oy 2236 TCTGAAAGAGATGATGAGATGTCTAGAAAGAACTCTGTGAAATTTAGTGAAGATG 2295
Db 2417 AATTAATAAGCTAGCAAGAAAGTGTAGAGAGCTTTGTAAAGATTTAGTGAAGAGTA 2476
Oy 2296 GAAGAACAGTGCAGAAAGATTAACCAAGAAATTTAGGACAAACCGGCACTGATCTAAAA 2355
Db 2477 AAAGAGCAATGCGGACATTAACAGAAATTTAAACAAACCGGCAAGATCTTGAAGAA 2536
Oy 2356 GATTTAAGCATATGAGAACTTAAGAAACGTGCAAGAGAAAGCAAGTCCAGT 2415
Db 2537 GATTTAAGCATATGAGAACTTAAGAAAGCAAGGCAAGAAAGCAAGTCCAGC 2596
Oy 2416 CTTGTTTGTCACTCATTAAGAAAGCAAGATATGATATCAAAAAGTAAATAGCAAAAC 2475
Db 2597 CTTGTTTGTCACTCATTAAGAAAGCAAGATATGATATCAAAAAGTAAATAGCAAAAC 2656
Oy 2476 AAGATTAAGATGCGGTTTCAACCGAATTCAGATTAACCAAGAAACATGTAATACTA 2535
Db 2657 GAAGATTAAGATGCGGTTTCAACCGAATTCAGATTAACCAAGAAACATGTAATACTA 2713
Oy 2536 CCGAGGAGATTAAGATGATCCGTAAAGAAATTAAGCTTAAGCTTAAGTGGCA 2595
Db 2714 CCGAGGAGATTAAGATGATCCGTAAAGAAATTAAGCTTAAGCTTAAGTGGCA 2773
Oy 2596 GCAGAGATTTGGAAGATATGATTTTGAAGAAAGATGATTAATTTGAATCAGAT 2655
Db 2774 GCAGAGATTTGGAAGATATGATTTTGAAGAAAGATGATTAATTTGAATCAGAT 2833
Oy 2656 TGCAGAAATTAAGAGATGCAAAAGCTTGAAGAAAGATGCAAAAGATTAATTAAGCT 2715
Db 2834 TGTGGAATTAAGAGATGCAAAAGCTTGAAGAAAGATGCAAAAGATTAATTAAGCT 2893
Oy 2716 TGTGCAATCTGAAGCTCTGAGAGTGAAGCCGCAAGAAACAGTGAAGAAATTAAGCAAG 2775
Db 2894 TGTGCAATCTGAAGCTCTGAGAGTGAAGCCGCAAGAAATTAAGTGAAGAAAGCAAG 2953
Oy 2776 ACAACTCAACCAACCAACCAACCGTTGCCATCCGAGGCAAGGCAAGTGAATCTCTTA 2835
Db 2954 ACAGCCCAACCAACCAACCAACCGTTGCCATCCGAGGCAAGGCAAGTGAATCTCTTA 3013
Oy 2836 CAGACCAAGACATGAGTTTACAGACATTCGACACACACACACACATCTTACATCA 2895
Db 3014 CAGACCAAGACATGAGTTTACAGACATTCGACACACACACACACATCTTACATCA 3073
Oy 2896 TCTACCATCATTAATAATTAATTTGACATCAACGAGGCAATGCAACCAAGTGT 2955
Db 3074 TCTACCATCATTAATAATTAATTTGACATCAACGAGGCAATGCAACCAAGTGT 3133
Oy 2956 ACAGCAGAGAGAAAGATGATGACAGAGAGCTGAAACCGAGTGAAGGCTGAGGATGAGT 3015
Db 3134 ACAGCAGG-----GATGATGACAGAGAGCTGAAACCGAGTGAAGGCTGAGGATGAGT 3187
Oy 3016 GGTGGAATGATGAGAGGAGGATGATGAGCAATGTTATTTGTTTATGATTTAG 3072
Db 3188 GGTGGAATGATGAGAGGAGGATGATGAGCAATGTTATTTGTTTATGATTTAG 3244

RESULT 10
AR438438
LOCUS AR438438 3081 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 9 from patent US 6664053.
ACCESSION AR438438
VERSION AR438438.1 GI:42663293
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 3081)
AUTHORS Kovacs,J.A., Huang,S., Mauer,H., Fischer,S.H., Gill,V.J. and Mei,Q.
TITLE Identification of a region of the major surface glycoprotein (MSG)
gene of human Pneumocystis carinii
JOURNAL Patent: US 6664053-A 9 16-DEC-2003;

FEATURES
Source Location/Qualifiers
1..3081
/organism="unknown"

ORIGIN
/mol_type="genomic DNA"

Query Match 52.9%; Score 1624.8; DB 6; Length 3081;

Best Local Similarity 72.8%; Pred. No. 1.2e-268;

Matches 2275; Conservative 0; Mismatches 752; Indels 99; Gaps 10;

```

Oy 1 ATGGCGCGGCGGTAGCGGCGAG-----GCAGCAGGACACAGAAATGACATTGAT 51
Db 1 ATGGCGCGGCGGTAGCGGCGAGCGTGGAAAGACATCAGGGCGTAGTGTAATATGATGT 60
Oy 52 GAGGAACTGTTTACCTTAACTTAAAGAGATGCACTAAGTACGACGAGATGCAAA 111
Db 61 GAAAGAAATCTTTGGCTTAAATGCA---GGAAAAAAATATATGATATATGATGCAAA 117
Oy 112 AAAAACTPAAAAAATATTCGCAAAATGACTGAAAGCAAACTPAAATATATGAAAGATA 171
Db 118 AAGAAATTAGAAAAATATTTGAAACATTTACGATGCAAGATTTAAACAGAAAAAGTT 177
Oy 172 CACAGAAAACTTAAAGTTTTCGAAAGATGAAAAAGCAGATACAAAATGCAAAAGACTG 231
Db 178 CACAAAAAATCTTAAGAGATTTTGTGAAATAAAAAAGCAGATTCAAAATGCAAAAGACTG 237
Oy 232 AAAGCCAAATTTGAGAAAAATGTACTACATCAAGAAAACTTAAAGAACAAATTAA 291
Db 238 AAAGAAAAATCTCAAAAAATGTACTGCAATCAAAAGAAAACTTACAGAACAACTCAA 297
Oy 292 AAAAAATTCAGATTATACGAGTAAAGATTGCAAAAGAAATGAACAACAAATGCTATT 351
Db 298 AAAAAAATTCAGATTATACGAGTAAAGATTGCAAAAGAAATGAACAACAATGCTATT 357
Oy 352 TTGGAGGAGATGTTCTCAAAAGAACTTAAAGATGATTGCAATCTTTGAAATAAGTGC 411
Db 358 TTGGAGGAGATGTTCTCAAGCGGAACTTAAAGATGATTGCAATCTTTGAAATAAGTGC 417
Oy 412 TATCAAAAAGAAAGTGATTAAGTTGCGGAAAGATTTCTTTAAGACACTTCGATGCA 471
Db 418 TATCAAAAAGAAAGTGATTAAGTTGCGGAAAGATTTCTTTAAGACACTTCGATGCA 477
Oy 472 CTTATGATCAGTATATGTAAGAAAAAATTAAGAGATTGGCCCTGTCATGGGAGG 531
Db 478 CTTATGATCAGTATATGTAAGAAAAAATTAAGAGATTGGCCCTGTCATGGGAGG 537
Oy 532 GAAAGTGAATGAATCAAACTTGTCTGAACCAAGAAAGACATGTAAGATATTTTAA 591
Db 538 GAAAGTGAATGAATCAAGAGCTTGTCTTATCAAAAAATGACGTGCAAAACATTTGTA 597
Oy 592 ATTGAAAAAGATTAAGAGTGCCTACTCTTAAACAGATGTTTACGACGACTAG---- 646
Db 598 TTAGAAAAAACAATAAAATGTATGCTCTTAAACAGATGTTTAAAGCAGCTTGAAG 657
Oy 647 -GAAATTTTAAAAAAGAAACATGCTGTAATTAATGCAACATGTAATTTTAACTTGA 705
Db 658 AAGAGTGAATTAAGAGAAATGTTTACCACTGCTTGAACGATGCTATTTTATGAGGG 717
Oy 706 AATTCGCGAG-----ACGACGATTAAT 729
Db 718 AATTCGAGAGATATATCAAAATGTAATTAATGATCCGAAAGATGTTATGAAATTTGCA 777
Oy 730 AATGTAATTAAGTGGAGAAATATGCCAAGAACAAATATTTGCTTATATGCGACAGGA 789
Db 738 GTGTGTAATCAATTTGGCAGTGAATGTGAAGAAATATAGATTATTTATACATCCGGGA 837
Oy 790 CCGGATTTGATCAACTAGCGCAGAGCTACAAATAGCAGAGATATAGGGCTGGAAGG 849
Db 838 TCCGATTTCAATCAACTAGCTAGCTAGCAGAGACATAGAGACTGGAAGG 897
Oy 850 TTTTATTAAGAGATGAGATGATTTTATTTATTTGAAAGATCATCTAAGAGATGCG 909
Db 898 CTTTATTAAGAGATGAGATGATTTTATTTGAAAGCTCTCTCTATGAGAGATGCA 957

```

```

Oy 910 ACAGCTTTGTGGCAATGTTGATTCAGAGATTCTAGTCTTAAAAAAGACGACAAAGAG 969
Db 958 ACTGCTTACTAGGGCGCTTTGTGATTCAAATCTTAGATCTTAAAGATCAAGTGGTAAAG-- 1015
Oy 970 AATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAATCCTCATGAACATGAGCTTTAGAA 1029
Db 1016 -AATGCGAAGAGCTTTTAAAGATTAATCTGTAAGAGTTTAAAGATCATGAATTTTGGGA 1074
Oy 1030 AGTTATGTAAGAAAAATGTTTAAATGATGTAAGGAAAGAAAAATGTAAGAAATTCGA 1089
Db 1075 GATTTTGTATCAAAATGATGCTGTCAAATGAAATGAAATGAAAGTGAAGAGTTGAG 1134
Oy 1090 AATGATATTAACAACCTGCAAAATTTTCACTTCAAAAGTCATATATATCGCTTTT 1149
Db 1135 AAGAGTTAGCAACAGTACTAAATCTTTTAAAAAATAAGAAATAAACACCTCTCT 1194
Oy 1150 GATCCAACAAAAAGAAATTAATGAATTTGATGAGGAAAGGTTGCCAACTTTCTTAGC 1209
Db 1195 -----GATCCGAGAAAGTCAATTCATGATGTAAGTTGACGACATTTCTTAGT 1242
Oy 1210 AACGAGATTTGCGAAATTTGAGTCTTATTTCTATTTTGAAAAAAATGTCAGAT 1269
Db 1243 GACAAATGCTGCAACAGGTTAGATGACAGCTGTTTATTT---AAAAAGTCAAGACCT 1299
Oy 1270 GGAGAAAAATGATGTAATAATTAAGCAACATGTTACAAAGAGACTTGATGCAAGG 1329
Db 1300 CTTACAAAGATTAATTAATCTGAAAGCAGATGTTATTAAGAAAGGCTTGAAGCACAA 1359
Oy 1330 GCAATATAAGTCTGCAAGAAATATGCGAGAGATGTTACATGTTCCAAACAAAAGCTGG 1389
Db 1360 GCTATATGAGATTTGCAAAAAAATGATGACGACCTGTTCTATGTTGACGCAAAAGATGG 1419
Oy 1390 CTTGAAAAATTTCAACAAAGATTAAGTATGTAAGAAATGTAAGAAAGAAAAAACA 1449
Db 1420 TTTAAGAAATCTATGAAAAAATATAGGAAGATGTTGCGAACTTAAACA----- 1470
Oy 1450 GGAAGTTCTCAACGATGAATTTATTTCTGTGTGACAGCCAGCAAAAGCAGCCCGG 1509
Db 1471 -----ACMACGATGAGTGTGTTTGTATGATGATGATCCACTTAAAGCGTCA 1521
Oy 1510 TTGCTTACACATGATCTTGCAGATGAAAACTATCTTTTACAGACAACTGGATCAAAAG 1569
Db 1522 ATACTTGAAGCTGATATTCACAGCAAGACAGCTCTTTTGGCGAAACAAATGGATCAAAAG 1581
Oy 1570 CGAGATTTCCGACAGATTAATAATTTGCAAGAAATTTGGGAGAAAGTGCACAAAGATTTAGA 1629
Db 1582 CGAGATTTCCAGACAGAAAGATTTGCAAGAAATTTAGGAAGAAAGTGAAGCTTTAGGG 1641
Oy 1630 GAGATTTCAAAAGAAATTAATCATGCGCATGTCATACATGAGACGAAATGCAATTCGCTTG 1689
Db 1642 AAGATTTCAAAATCAATTAAGTGGCCATGTCATACGCTTAAACAAACAGATGATGCTTG 1701
Oy 1690 GGAATCAAGAAATTTTAAAGCAGATTTTATTTGATGAAACAAAGATTAATTTTAAAGAC 1749
Db 1702 GGAATCAAGAAATTTTAAAGCAGATTTTATTTGATGAAACAAAGATTAATTTTAAAGAC 1761
Oy 1750 CAAGAAATTTGTAATAATTAATTAATAAGAAAGTGAATAATGCTTGAAGAGAGAT 1809
Db 1762 CATGAAATCTGTAAGAAATTTTAAAGAAAGTGAATAATGCTTGAAGAGAGAT 1821
Oy 1810 GACGTTTCTCTTTTGTATGTGTCTTCCAAACGCTACGTGAGCTGATGTAATAAGAC 1869
Db 1822 GATGTTTCTCTTTTGTATGTGTCTTCCAAACGCTACGTGAGCTGATGTAATAAGAT 1881
Oy 1870 GTGAAGCAGATGTAATTAATAAGAAATTAATAAGCTTCAATATTAATGAATTT 1929
Db 1882 GTGAAGCAGATGTAATTAATAAGAAATTAATAAGCTTCAATATTAATGAATTT 1941
Oy 1930 CTTGAAATTAATAATAATAATAACAACATGGAAGAAATTTGCTTGGCATGCG 1989
Db 1942 CTTAATAATAATAATAATAATAACAACATGGAAGAAATTTTGTGCTTCAATGCAACCA 2001

```

QY 1990 TATTGATAGATTTTTCACCTTAATGTCAGAGCTTACGAAAGAAATAGT---TGTAACA 2046
 DB 2002 TACTGCGATATATTTTACCCAAATTGCTCTGATCTTAAGAAAGAAAACTTCTGTCGA 2061
 QY 2047 AAAATCAAGAACATCGTAGCCGCTCTATTAAGAAAGCCCTTGAAAGATGCTCTCAA 2106
 DB 2062 AATCTTAAAAAATATGCGAACCATCTTACAAAAGAAAGGTTTAAAGATGCTCTTAAA 2121
 QY 2107 GTAGAGCTTCAAGGAAATTTGATGATTAATCTAAATGTAACCTGCACTTGAAGAAAT 2166
 DB 2122 GTAGAGCTTCAAGGAAATTTGATGATTAATGTAATGTAATGCACTTGAAGAAAT 2181
 QY 2167 TGTAAGTAGCGGGAACGTAATATATGCTCAATCACTGCTTATGCAAGTCAACACC 2226
 DB 2182 TGCAATATTTGAAAAATGTAAGTATCATCAATCAACAGTTTATGTAAGATTAATCC 2241
 QY 2227 AAGATTAATCTTGAAAGAGTGTAGAGATCTTAAAGAAAGAACTTCTGTGAAATTAAGT 2286
 DB 2242 GAAAGTAAAACTAAAAAGACCGATATGAAAGTAAAGAAAGCTTTGCTAAAAATTAGTG 2301
 QY 2287 AAAAGAGTGAAGAACATGCAAGCATTAACCAAGAAATTAGCAACCGGACGCTGAT 2346
 DB 2302 GAAAGAGTGAAGAACATGCAAGCATTAACCAAGAAATTAGCAAGCAAGAAAGAAAC 2361
 QY 2347 CTAAGAAAAAGATTAATAGACATATGAGCACTTAAGAAAGCTGCAAGAAAGCAATGAA 2406
 DB 2362 CTAAGAAAAAGATTAATAGACATATGAGCACTTAAGAAAGCTGCAAGAAAGCAATGAA 2421
 QY 2407 AAGTCCAGCTTTGTTGTCACCTCATTAAGAAAAAGAAAGTATGATCAAAAAGTAA 2466
 DB 2422 AATTCATATCTTTGTTTATCATCTTCCTTAAGAAAGTAAATTAATCATCAAAAATAGT 2481
 QY 2467 AGCAAAAAAGATTAAGAAATGCGCTTCAAGCAAGCTTCAAGAAAGTATCAACAAACATG 2526
 DB 2482 AGCAAAAAAGATTAAGAAATGCGCTTCAAGCAAGCTTCAAGAAAGTATCAACAAACATG 2541
 QY 2527 AAAATTAATCAAGAGAGAGTAAAGATGATCCGTAAGCAAGATTAAGAAAGTAAAGCAT 2586
 DB 2542 AAAATTAATCAAGAGAGAGTAAAGATGATCCGTAAGCAAGATTAAGAAAGTAAAGCAT 2601
 QY 2587 GATTGGCAGAGAGATATTTGAAAGATATGATGATTTGAAAGAAAGTAAATTAATG 2646
 DB 2602 GATTGGCAGAGAGATATTTGAAAGATATGATGATTTGAAAGAAAGTAAATTAATG 2661
 QY 2647 GAATCAAGTTCCAAATTAAGAGATGTCAAAGCTTAAAGAAAGTATGCAAAAAGTT 2706
 DB 2662 GAATCAAGTTCCAAATTAAGAGATGTCAAAGCTTAAAGAAAGTATGCAAAAAGTT 2721
 QY 2707 AATAAGGCTTGTGCAATCTGAAGCCTCTGAGGTGAAGCCGCAAGAAAGTGAAGAA 2766
 DB 2722 CAAAGAGTATGTTGCAATTAAGAACTGAAGAGTGAAGCCGCAAGAAAGTGAAGAA 2781
 QY 2767 AGTCAAGAGCACTCAACAAACAAACAAACCGTTGCCGATCCGAAAGCAACGGAATG 2826
 DB 2782 AGTCAAGAGCACTCAACAAACAAACAAACCGTTGCCGATCCGAAAGCAACGGAATG 2841
 QY 2827 AAAATCTTACAGCAAGCAAGCATGAGTTACAGAGATGACACACACACAGCAGCT 2886
 DB 2842 AAAATCTTACAGCAAGCAAGCATGAGTTACAGAGATGACACACACACAGCAGCT 2901
 QY 2887 ACTATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2946
 DB 2902 ACTATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2961
 QY 2947 ACCAAGTGAAG 3006
 DB 2962 ACCAAGTGAAG 3015
 QY 3007 AGATAGATGAGTGAATGATGATGAGGGGGGATGATGATGATGATGATGATGATGATG 3066
 DB 3016 AAGATGATGAGTGAATGATGATGAGGGGGGATGATGATGATGATGATGATGATGATG 3075
 QY 3067 ATTAG 3072

Db 3076 ATTAG 3081
 RESULT 11
 AF033212
 LOCUS
 DEFINITION
 PNEUMOCYTIS CARINII F. SP. HOMINIS CLONE HUMSG32 MAJOR SURFACE
 GLYCOPROTEIN (MSG) GENE, PARTIAL CDS.
 ACCESSION
 AF033212
 VERSION
 AF033212.1
 GI:3560520
 KEYWORDS
 PNEUMOCYTIS JIROVECI
 PNEUMOCYTIS JIROVECI
 BAKTERIA; FUNGI; ASCOMYCOTA; PNEUMOCYTIDOMYCETES;
 PNEUMOCYTIDACEAE; PNEUMOCYTIS.
 REFERENCE
 1 (bases 1 to 3080)
 Garbe, T.R. and Stringer, J.R.
 Molecular characterization of clustered variants of genes encoding
 major surface antigens of human *Pneumocystis carinii*.
 Infect. Immun. 62 (8), 3092-3101 (1994).
 JOURNAL
 MEDLINE
 PUBMED
 94314421
 7518806
 2 (bases 1 to 3080)
 Mei, Q., Turner, R.E., Sorial, V., Klivington, D., Angus, C.W. and
 Kovacs, J.A.
 Characterization of major surface glycoprotein genes of human
Pneumocystis carinii and high-level expression of a conserved
 region.
 Infect. Immun. 66 (9), 4268-4273 (1998).
 JOURNAL
 MEDLINE
 PUBMED
 98380374
 9712777
 3 (bases 1 to 3080)
 Mei, Q., Turner, R., Sorial, V., Klivington, D., Angus, C.W. and
 Kovacs, J.A.
 Direct Submision
 Submitted (07-NOV-1997) Critical Care Medicine Dept., National
 Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD
 20892-1662, USA
 FEATURES
 source
 1..3080
 /organism="Pneumocystis jirovecii"
 /mol_type="genomic DNA"
 /specific_host="Homo sapiens"
 /sub_species="hominis"
 /db_xref="taxon:42068"
 /clone="HUMSG32"
 /note="derived from HIV-infected human with *P. carinii*
 pneumonia"
 <1..3029
 /gene="MSG"
 <1..3029
 /gene="MSG"
 /gene="MSG"
 /note="surface antigen"
 /product="major surface glycoprotein"
 /protein_id="AAC34975.1"
 /db_xref="GI:3560521"
 /translation="AAAVKROAAKASGASVYDGEIILALIAKKYNDNECKELEK
 CKITLTDALPEKVKHKKLKEFCENKSKDSCKEIKELTQKTAIKGLTEAIKKNS
 DLTDDEKENOQCLPLEGACPAELKBDCTLBSDFYLSQAPLDEKCNLKAQYRGLE
 NRTCEGKLKEVCEIELSDESDDELTKLCLYQMTCTFVLREKOKCNALKODVNALEK
 KDELRGKCLPLLRVCYPRNCEIDISCKNSSEECVYLVLCOTLAKECENKIITYH
 PGSDFNPTKSKPTAEDIGIBELYLKKAABEGVNIQKRPVADATLALLLIONDPRKO
 VGRCEVLDKNCLEIKSHSILGDFQNVAGQNEIKCKELEKLNASTKILIEKIK
 NKHLSGEVIIPWYKLTPLSDNDCTLSBDFYLSQAPLDEKCNLKAQYRGLE
 AQANEAOLKQKMYGLFYSGKEMFKLLEKIMECESEIKTSDLEFLICIDPLKAVRII
 AADIOARAVFLRKOLDKDPEDPKDCEKEIGKCEALGXDSNOIKMFCITLKQCCDL
 GTTEILKOVILDEHKOTLRHENTCTYLKCKCHKWASRRGDRSFVQVONATCKLIV
 DDVDRCEVBEKMKQASDINSLKNOIKTESAANIPSPHRYCDRLPLNCPDCKKK
 TPCQNLKRYCEPFRKRVLDALKEVLEQSLSNKCESALERTCTILKQVDSINS
 LCKDNTESKTKTQDNEVRKQLCLKLVEVEQOQKMLPALEHEBKDLDFETEKIK
 KQAEKTMKSNLVLSPVKDEBNNTSKSSKDKDKNTVSNGLDITTEHMKILIRGVKDV

SVTSEAKAFDVAIVFGRYLDKERCKLSDCKVKKDCKDLBSVCKIGVCSKLK
PLKVPHEIVTESITTTTTTTTTTVDTPATECKSLQTDWTITQSTHTSTITSTI
ISKILTSIRCKPKCTGDAEDVKEBGLKMSG"
3029
/gene="MSG"
/note="putative PCR artifact leads to premature stop
codon"

ORIGIN

Query Match 52.9%; Score 1623.8; DB 8; Length 3080;
Best Local Similarity 72.8%; Pred. No. 1.8e-268;
Matches 2274; Conservative 0; Mismatches 752; Indels 99; Gaps 10;

2 TGGCCGCGGCGGTCAAGCGGCAG-----GCAAGCGGACACAGAAATGCAATTGATG 52
1 TGGCCGCGGCGGTCAAGCGGCGGTGCAAAAGCATCAGGGCGTGTATGATGATGCTG 60
53 AGGAACATGTTTACTTTAATTTCAAGAGATGGACTAGTAGCAGGAAATGCAAAA 112
61 AAGAAATTCCTTTGCTTTAATTGCA--GGAAAAAATATATATATATGATGCAAAA 117
113 AAAAACTAAAAAATATGCGCAAGATTGACTGAAGCAAACTAATATAGAACAGTAC 172
118 AAGATTGAAAAAATTTGTAGACATTAAACGATGACGAATTTAAACCAAGAAAGTTG 177
173 ACAGAAAACTTAAAGTTTTCGCAAGATGAAAAACGATACAAAATGCAAAAGACTGA 232
178 ACMAAAACCTTAAAGAGTTTGTGAAAAATMAAAAGCATTCMAAAATGCAAAAGACTGA 237
223 AAGCAATATGAGAAAAATGACTACATCAATCAAGAAAACTTAAAGAGCAATTAATA 292
228 AAGAAAACTCTCAAAAATGATCTGCATCAATCAAGAAAACTTACAGAGAAATCAAAA 297
293 AAAAAATTCAGATTATTAACGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 352
298 AAAAAATTCAGATTATTAACGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 357
353 TGGAGGAGTATGTTCAAAAAGACTTAAAGATGATTCATATCTTTGAGAAATAGTGT 412
358 TGGAGGAGTATGTTCAAAAAGACTTAAAGATGATTCATATCTTTGAGAAATAGTGT 417
413 ATCAAAAAGAGTGTAAAGTTGCGGAAGAGTTCTTTAAGACACTTGTAGCATG 472
418 ATCAAAAAGAGTGTAAAGTTGCGGAAGAGTTCTTTAAGACACTTGTAGCATG 477
473 TTAATGATCACTCATATGTGAAAAAATCTTAAAGAGATTTGCCCTGTCAATGCGGAGG 532
478 TTAATGATCACTCATATGTGAAAAAATCTTAAAGAGATTTGCCCTGTCAATGCGGAGG 537
533 AAGATGATGATTAACAACTTGTGTCTGACCCAGAAAGACATGTAAGAAATATTTTA 592
538 AAGATGATGATTAACAACTTGTGTCTGACCCAGAAAGATGACGTAAGAAATATTTTA 597
593 TTGAAGAAAGATTAAGAGTGTGCTTAAACAGATGTTTCAAGACAGCTAG----- 646
598 TTGAAGAAAGATTAAGAGTGTGCTTAAACAGATGTTTCAAGACAGCTTGAAGAA 657
647 GAAGTTTAAAAAAGAAATCATGCTTGAATTAATGCAACATGCTATTTTTCATTTGAA 706
658 AAGATGATGATTAACAACTTGTGTCTGACCCAGAAAGACATGTAAGAAATATTTTA 717
707 ATTGCGGAG-----ACGACGATATTAATTA 730
718 ATTGGAAGATATATCAAAATGTATTAATATCAATCCGAGAGCTGTATGATTTTTCGAG 777
721 AATGATTAATTTGGAGAAAAATGCGCAAGAACTTGTCTTATATGCAACAGAGAC 790
778 TGTGTGATATCATTTGGAGTGAATGTGAAGAAATTAAGATTAATTAATCAATCCGAGAT 837
791 CCGATTTTGAATCAACTAGCCAGAGGCTACATATGACAGAGATATAGGGCTGGAAGAT 850
838 CCGATTTTGAATCAACTAGCTAAGCCCTACTGTAGCAGAGACATATGAGACTGGAAGAC 897

851 TTTTAAAGAGTGAAGAGATGAGTGTATTTTATTGAAAAAGATCATCTAGAGATGCGA 910
898 TTTTAAAGAGTGAAGAGATGAGTGTATTTTATTGAAAAAGATCATCTAGAGATGCGA 957
911 CAGCTTTGTTGGCATTTGTTGATTCAGATTTCTAGCTTTTAAAAAAGAGACAGCAAGAGA 970
958 CTGCTCTACTGCGCTTTTGTATTTTAAATCTAGATCTTAAAGATGATGATGATGATG 1014
971 AATCGAAGAGGCTTTCAAAAAAGCTGCAAAAATCTCATGAAATGATGAGCTTTAGAA 1030
1015 AATCGAAGAGGCTTTCAAAAAAGCTGCAAAAATCTCATGAAATGATGATGATGATG 1074
1031 GTTATGTAAGAAAAATGTTTAAATGATGATGATGATGATGATGATGATGATGATG 1090
1075 ATTTTGTAAATCAAAATGATGATGATGATGATGATGATGATGATGATGATGATG 1134
1091 ATGATATTAACAAATCTGCAAAATTTTCACTTCAAAAGTCACTAATATGCTCTTTTG 1150
1135 AGGATTAAGCAACATGATTAATTTCTTTTGAAGAAAAATTAAGATTAACATCTCT 1193
1151 ATCGAAGAGGCTTTCAAAAAAGCTGCAAAAATCTCATGAAATGATGAGCTTTAGAA 1210
1194 -----GATTCGAGAGAGTCAATTCATGATGATGATGATGATGATGATGATG 1242
1211 ACAGAGATGATGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATG 1270
1243 ACAATGATGCAAGGTTAGATGATGATGATGATGATGATGATGATGATGATGATG 1299
1271 GAGAAATGCAATGATTAATTAAGAGCAATGATTAACAAAGAGACTGATGACAGG 1330
1300 TTGACAAAGATGATTAATTAAGAGCAATGATTAACAAAGAGACTGATGACAGG 1359
1331 CAAATTAAGTGTGCAAGAAATTAAGAGCAATGATTAACAAAGAGACTGATGACAGG 1390
1360 CTAATTAAGGCTGCAAGAAATTAAGAGCAATGATTAACAAAGAGACTGATGACAGG 1419
1391 TTGAAGAGTTCACAAAGATTAAGAGCAATGATTAACAAAGAGACTGATGACAGG 1450
1420 TTAAGAAATCTAGTGAAGAAATTAAGAGCAATGATTAACAAAGAGACTGATGACAGG 1469
1451 GAAGTTTCTCAAGATGATTAATTAAGAGCAATGATTAACAAAGAGACTGATGACAGG 1510
1470 -----ACAAGGATGATGATGATGATGATGATGATGATGATGATGATG 1521
1511 TGTTCACATGATCTTTCGATGAAATTAATCTTTTTCGACAACTGATGATGATG 1570
1522 TACTTGCAGTATGATTCACAAAGAGAGCTTTTTCGAGAACTTTCGATGATGATG 1581
1571 GAGATTTCCGACAGATTAATTAAGAGCAATGATTAACAAAGAGACTTTCGATGATG 1630
1582 GAGATTTCCGACAGATTAATTAAGAGCAATGATTAACAAAGAGACTTTCGATGATG 1641
1631 AGGATTTCAAGAAATTAATTAAGAGCAATGATTAACAAAGAGACTTTCGATGATG 1690
1642 AGGATTTCAAGAAATTAATTAAGAGCAATGATTAACAAAGAGACTTTCGATGATG 1701
1691 GAGCTTACAGAAATTTTAAAGAGGTTTATGATGAAACACAAAGATCTTTGAAAGAC 1750
1702 GAGCTTACAGAAATTTTAAAGAGGTTTATGATGAAACACAAAGATCTTTGAAAGAC 1761
1751 AAGAAAGTGTGTAATTAATTAAGAGCAATGATTAACAAAGAGACTTTCGATGATG 1810
1762 ATGAAAGCTGTAAGAAATTAATTAAGAGCAATGATTAACAAAGAGACTTTCGATGATG 1821
1811 ACCGTTTCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1870
1822 ATGTTTCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1881
1871 TGAAGACAGGTGTAAGATTAATTAAGAGCAATGATTAACAAAGAGACTTTCGATGATG 1930
1882 TGAAGACAGGTGTAAGATTAATTAAGAGCAATGATTAACAAAGAGACTTTCGATGATG 1941
1931 TTGAAGATTAATTAATTAAGAGCAATGATTAACAAAGAGACTTTCGATGATGATG 1990

Db	1942	TTAAAAATTAACAAATATAAAAACAGATATAGACGCAAAATATTTGTCCCTCATGCAACCAT	2001
Qy	1991	ATTGCAATAGATTTTTCACCTTAATTGTCAGGCTCTTACGAAGAGATAGT--TGTAACA	2047
Db	2002	ACTCGATAGATTTTTCACCAATATGTCGTATCTTAAGAAAGAAAACTTCTGTCAAA	2061
Qy	2048	AAATCAAGAGACATCGTGAAGCCGTTCTATTAAGAAGAGCCCTGGAGAGTCTCTCAAG	2107
Db	2062	ATCTTAATAAAATATATGCGAACCATTTCTACAAAGAGAGGTTTGTGAAGATCTCTTAAG	2121
Qy	2108	TAGAGCTTCAAGAAAAATGACTGTGTAACTTAAATGTGAACCTGCACTTGAAGATAT	2167
Db	2122	TAGAGCTTCAAGGAAATTTAACTAATATAAAATAAATGTGAATCTGCACTTGAAGATAT	2181
Qy	2168	GTACAGTAGCGGGAACGTAAATATATGCGTCAATCAGTGGCTTATGCAAGCTAACCA	2227
Db	2182	GCAACATATTTGAATAATGTAAGTATTCATCAATCAAGTTTATGTAAAGATATACG	2241
Qy	2228	AGGATAACTCTGGAAGAAGTATGAGATGCTGTAAGAGAACTCTGTGAGAAATTAATGA	2287
Db	2242	AAAGTAAAACTTAAAGAAAGCCGATATAGAACTTGAAGAAAGACCTTGTCTTAAATTAAGTG	2301
Qy	2288	AAGAAGTGAAGAACAGTGCAGAACATTAACAAGATTAGAGCAACCGGCACTGATC	2347
Db	2302	AAGAGGTGAACACACATATGTAAATGTTATCAAGCAAGTATGAGATGAGAAAAAGACC	2361
Qy	2348	TAAAAAAAGATTTATAGACATATGAGAACTTAAAGAACTGTGACAGAGAACATAGACA	2407
Db	2362	TAAAGATGATTTTGAACAATTTGAAAACTTAAABAAACAGCAGAGAAAAACATAGATA	2421
Qy	2408	AGTCAGTCTTGTGTTGTCACTCATTTAAGAAAAAGAAAGTATGTATCAAAAAATATA	2467
Db	2422	AATCAATCTTGTGTTTATCATTTGTTAAGAAAGATGAATAATATACATCGAAAAATAGTA	2481
Qy	2468	GCAAAAAACAAGATTAAGAAATGCCGTTTCAACCGACTTCAAGATATCCAACAAACATGTGA	2527
Db	2482	GCAAAAGACAAAGATTAAGAAATCCGTTTCAACCGACTTCAAGATATCCAACAAACATAGTA	2541
Qy	2528	AAATATCTACGAGGGGAGTTAAAGATGTATCCGTAAACAGAAATTGAAGCTTAAAGCATTTG	2587
Db	2542	AAATATCTACGAGGGGAGTTAAAGATGTATCCGTAAACAGAAATCTGAAGCTTAAAGCATTTG	2601
Qy	2588	ATTGCGCAGCAAGATATTTGGAAGATATGTAGATTGGAAGAAAGATGTATAAATTTGG	2647
Db	2602	ATTGCGTAGCAGAAAGTATTTGGAAGATATCTAGACTTGAAGAAAGATGTATAAATTTGG	2661
Qy	2648	AATCAGATTCGAGAATTAAGAGATTTGCCAAGACTTGAAGAAAGTATGCAAAAAAGATTA	2707
Db	2662	AATCAGATTCGAGAAGTTAAAGAGATTGCCAAGATTTGAAGAGATATGTGGAAGAGATC	2721
Qy	2708	ATAAGGCTTGTGCGAATCTGAAGCCTCTGAGGTGAACCGCAGAAACAGTGCAGAGAA	2765
Db	2722	AAGAGATATGTTCGAAATTTAAACCACTGAAAGTGAACCGCAGAAACAGTGCAGAGAA	2781
Qy	2768	GTACACGACAACTTACAAACAACAACAACCGTTGCCATCCGAAGGCAACGGAATGCA	2827
Db	2782	GCAACAGACGACCAACGACGACAAACAACAACCGTTATCGATCCGAAGGCAACAGATGCA	2841
Qy	2828	AATCTTTACAGACAAACAGACACATGGGTTTACACAGACATCGACACACAAAGCAGTCTTA	2887
Db	2842	AATCTTTACAGACAAACAGATACATGTGATTAACAAGACTTGCACATATCCAGCAGTCTTA	2901
Qy	2888	CTATCACACTTACCATTCATCAAAAAATTAACATTGACATCAACGAGGGGAGTGCAAACCA	2944
Db	2902	CCATCACACTTACCAATTCATCTAATAAATAAATCTCACTCACTCAACAGGCGTTGCAAACCA	2961
Qy	2948	CCAAGTGTACACAGAGAGAGAAAGATGTATGCAGAGAGCTGGAACCGAGTGAAGGGCTGA	3007
Db	2962	CCAAGTGTACAGACAG-----GAGTATGCAAGAGAGCTGAAGCCGAGTGAAGGGAATTGA	3015
Qy	3008	GGAATGATGGGTGGAATGTATGAGGGGGGTGATGTAGCAATGTTATTTTCGTTCAATGA	3067

Db	3016	AGATGACTGGGTAAACGTANTGAGGGGGGTGATAGTACGAAATGGTATTATTCGTTATCA	30715
Qy	3068	TTTAG 3072 	
Db	3076	TTTAG 3080	
RESULT 12			
LOCUS	AF038556	12792 bp	DNA linear PLN 10-SEP-1998
DEFINITION	Pneumocystis carinii f. sp. hominis variant regions of major surface glycoprotein (meg1, meg3, meg4) genes, partial cds.		
ACCESSION	AF038556		
KEYWORDS	AF038556.1 GI:3560524		
ORGANISM	Pneumocystis jirovecii Pneumocystis jirovecii Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;		
REFERENCE	1 (bases 1 to 12792) Met, O., Turner, R.E., Social, V., Kilmington, D., Angus, C.W. and Kovacs, J.A. Characterization of major surface glycoprotein genes of human Pneumocystis carinii and high-level expression of a conserved region		
AUTHORS	Met, O., Turner, R., Social, V., Kilmington, D., Angus, C.W. and Kovacs, J.A. Direct Submission Submitted (12-DEC-1997) CCMD, NIH, Building 10, Room 7D43, MSC 1662, Bethesda, MD 20892-1662, USA Location/Qualifiers		
TITLE	1. 12792		
JOURNAL	/organism="Pneumocystis jirovecii"		
MEDLINE	/mol_type="genomic DNA"		
PUBMED	/specific_host="Homo sapiens"		
REFERENCE	/sub_species="hominis"		
AUTHORS	/db_xref="taxon:42068"		
FEATURES	<668..3712		
SOURCE	/gene="meg1"		
	/note="multicopy gene family is expressed by gene rearrangement"		
	/note="glycoprotein A, expressed from a single putative leader sequence located upstream of coding region; variant region is in frame with the leader sequence"		
	/codon_start=1		
	/product="variant region of major surface glycoprotein"		
	/protein_id="AAC34980.1"		
	/db_xref="GI:3560525"		
	/translation="VAAVAKQYTGASGVDEEVRLLALIKEDSKDDKCEBELGKHCKEISEANLTPPEVHEKLDKFCDSKKRDKCKELKXNVEKCDPTELEBELVKEKASNLKNDCEKNEOQCLFLEBAGCDLTKNCNDLRKCYONKEDYAKAEVTLRIIGKDKPKDKSCENKLEVYCOELSGMSDEMLKLCFDPQNTCDNLVKETQOCSEFNLKLEIKTIKPEBOLKKKCPILYEBCTFYDESGCNSLKCELEKKCOEKNITLYTSYGPDIEPETTLAEEDVDSGIYRKAAEBSTYQKPLPAPATLVAVLIDPSLTORTMKCECKTILIDCKNAKHEHDIIGLCEEDYNANKDKOKKCEBSTDIEETCKFFISKLTIHFFGDGNKNQDILKGNLSTFLSNKDCITLSEYCLYEKSCRSBTACKNIRAAQYKGLDTLANEVLQKEMRGMHSGNKTMLSGFOKLLIECVKVKKNENGVPSNELPVLGVQSPKVALLSHLDIKRTIFLODILNRKRDPIVKECEBELLKKECALRKDSKIEWPCTLSQNCQQLBNAKKELLLENHNDILKNQNCQGLLEKKNENMRNRNRPSSILCALONTRCIMVEDVAKNCKIFEKRIKTYGSDISKTKIBELGTCTPIYHPRCHRGPNRPLEKKKCDPEYCKRKYRQRLDENALIFEFRGHDKKKNCKTNLDKYCTLMDQYNTKTLKGFNSSTNNKTRFDLCEKLVORVEKCOGLSKELERKANDLEEHQYEDKQYKOTAKAMESTNLVASTKSTDNKTEKGVSTPSVQDVIHPLVYKNEKVOTEKAKAFDVLALFSLVLELTSGHKLKDCDFPEKCECKQCKQCEIKELCIKPLKVPKPEIKTIVETNITTTVETVVEAKTYGDEGKESLSTTDVWVKTSTHSTSTSTTSVYSRIILTSRCKCKPIKCTTDBEADGAEVKSBSGLRMSGMSVMRGVLLAMISIMI"		
misc_feature	4230..7225		

gene		/note="encodes a major surface glycoprotein but due to a frame shift has a premature termination; correction of the frame shift results in a complete open reading frame for msg2"	
		<7735..10743	
		/gene="msg3"	
		/note="multicopy gene family is expressed by gene rearrangement"	
		<7735..10743	
		/gene="msg3"	
		/note="glycoprotein A; expressed from a single putative leader sequence located upstream of coding region; variant region is in frame with the leader sequence"	
		/codon_start=1	
		/product="variant region of major surface glycoprotein"	
		/protein_id="AAC34981.1"	
		/db_xref="GI:3560526"	
		/translation="VARAVKRRRAAONSVEEYLTLILLENENNDKCKRKEVCEVKKVTPKPKKLEBKLDGICKDDKTBACKESETVYKACSPQELDKAVKASTLEBDDCKNRKQCCFLFEGACPTFLKDKCKNELRKCYQKRRDVAEALLRVKGLDKNTKCKNLKGVCOEFNKSEDELIKCLUDEKTCGDVSKREYCKPKEGIDVLVIGKEKTLDEKLEKLEBCEYGCNCEPDQPCFPAKCCOENLVAAPOSHDPDKLIRLAELEIDELKLYEAVKGIHIGRPSIKDEYVALLALSLKSAQVTPKQCCDVIKKCCGNFKHEIILKDKSNKTIITNDPKCEKECELNELTRILTVSRLEKPAVNEIIGMHMATHFRGERTCULSDCFYKLSQAPLKPCHNNTKACRYKLEPAVNAALQDKRGLOGSNRTWELTLQKNLVYKCEKTKGSDELPIVCMPIPTALTVSTLDRRAVALQEHLENGRDPTEPKDCELEKCEVLAGKSDRELKYSYTLKQHCNRLSLIHELSEELQENKYLSEIKNEBKAKRCEKMPFRBNKKFSPASDLELCKITRVNESCMIILGHHETMVLNBIANKBEKICEKMPAPYCKRYQOENCEGLKNGGODGCKLNRKCFPELKEALENKVELKGLSLNNGECNNTNITQTLKRNENLETLCKENKTSADIVREBELKIKIRLEKCKSLKDBELEPVKVELKEKCKYKILPEAKMEANILISAKPQDNMNNKSNKSSDTPKESKGTGTFKLVRRNAKHVTEKELAADVLVARADVLBELKEICNHLSCGFKKEDCEDCPCKIQGISLTPLEKVPHEIVTKNITTTTTTTTTTIKDAKATDCHSLQTDITWTKTSTSTSTSTSTSTSTSTSTSTSTSTRCKPTCTTGEBDEAGDVPSBGLRMSGSVWRGVILAMTISFMI"	
		<11268..>112792	
		/gene="msg4"	
		/note="multicopy gene family is expressed by gene rearrangement"	
		<11268..>112792	
		/gene="msg4"	
		/note="glycoprotein A; expressed from a single putative leader sequence located upstream of coding region; variant region is in frame with the leader sequence"	
		/codon_start=1	
		/product="variant region of major surface glycoprotein"	
		/protein_id="AAC34982.1"	
		/db_xref="GI:3560527"	
		/translation="VARAVKRRRAAGTQNSIDSEHVALILKEDGSEQCKKLKRYCQELTEALNIEOVHKHLKGFCEBQKADYTKELKAIIEKKCTIIEKLEALRKNSDLTDLDEKENEQOCLFLEGVCSKELKDDCMTLRNKCQKRRDVAEALLRVKGLSDSVLEKLEKICEIPVMGRESDDELTNLCINQETCKNLIIEKDKCGTLKIDVSAALGSFKCEKCELLBOCYFVIGNCGDDIIKCIEELGKCOBONIAMPPEGDPDPTBPATIAEDGLEPFRKKVEDGAVFISGNHLRATNALIIDSSLKXKDKKEKCELOJSCCKNPNHEHLEBESLCKKNGLSNDGTYKCEBLQNDINKTKITSKYTRNRLDPTPKGNNEIVMGHGLPTLSNEDCAULBSYCFYFEKPCPDGNACKNIRATCYKRGDARANKVLOENMRGMLHGSNKSWMLEKFODELVVCERLKKENKSPSNDDELFLQVPAKARLLTHD"	
ORIGIN			
Query Match	48.8%; Score 1498.4; DB 8; Length 12792;		
Best Local Similarity	99.0%; Pred. No. 4.5e-247;		
Matches 1508; Conservative	0; Mismatches 16; Indels 0; Gaps 0;		
Db	11269	TGGCGCGGGCGGTGAACGGGACAGCCAGGACACAGATATGATGAGGAACATG 11328	
Qy	62	TTTTAGCTTTAATCTTAAAGGAAGATGACTAAGTGAGCAGAAATCAAAAAAATCTAA 121	
Db	11329	TTTTAGCTTTAATCTTAAAGGAAGATGACTAAGTGAGCAGAAATCAAAAAAATCTAA 11388	
Qy	122	AAAAATATTTGCCAAGAAATTGACTGAAGCAAACTAAATATATAGAACATGACAGAAAC 181	

Db	11389	AAAAATTTGCCAAGATTGACTGAAGCAAAACTAAATATAGAACAGTACACAAAAAC	11444
Qy	182	TTAAAGTTTTTCGAAGATGAAAAAGCAGATACAAATGCAAAAGACTGAAGCCAAT	241
Db	11449	TTAAAGTTTTTCGAAGATGAAAAAGCAGATACAAATATAGAACAGTACACAAAAAC	11508
Qy	242	TTGAGAAAAATGTACTACATCAATCAAGAAAACTTTAAAGACGAAATTTAAAAAAATTC	301
Db	11509	TTGAGAAAAATGTACTACATCAATCAAGAAAACTTTAAAGACGAAATTTAAAAAAATTC	11566
Qy	302	AGATTATATACGATTTAAGATTTGCAAAAGAGATGAACAACATGCTATTTTGGAGGAG	361
Db	11569	CAGATTTAACGGATTAAAGATTTGCAAAAGAGATGAACAACATGCTATTTTGGAGGAG	11622
Qy	362	TATGTTCAAAAAGACTTAAAGATGATTCATATCTTTGAGAAATAAAGTCTATCAAAAGA	421
Db	11629	TATGTTCAAAAAGACTTAAAGATGATTCATATCTTTGAGAAATAAAGTCTATCAAAAGA	11688
Qy	422	AACGTATAAAGTTGGCGGAAGAGTTCTTTTAAAGACACTTCGTACCGATCTTAATGAT	481
Db	11689	AACGTATAAAGTTGGCGGAAGAGTTCTTTTAAAGACACTTCGTACCGATCTTAATGAT	11746
Qy	482	CAGTCATATGTGAAAAAACTTAAAGAGATTTGCCCTGTCATGGGAGGAGAAAGTGAATG	541
Db	11749	CAGTCATATGTGAAAAAACTTAAAGAGATTTGCCCTGTCATGGGAGGAGAAAGTGAATG	11806
Qy	542	AGTTAACAAACTTGCTGTCGACACAGAAAGACATGTAAAGATATTTTAAATTTGAAAAAG	601
Db	11809	AGTTAACAAACTTGCTGTCGACACAGAAAGACATGTAAAGATATTTTAAATTTGAAAAAG	11866
Qy	602	ATTAAGAGTGGCGTACTCTTAAAAACGATGTTTACAGACACTAGAAAGTTTAAAAAG	661
Db	11869	ATTAAGAGATGGGTACTCTTAAAAACGATGTTTACAGACACTAGAAAGTTTAAAAAG	11922
Qy	662	AAACATGCTTGAATTAAGTGAACATGCTATTTTAAATTTGGAATTTGGAGAGAGAG	721
Db	11929	AAACATGCTTGAATTAAGTGAACATGCTATTTTAAATTTGGAATTTGGAGAGAGAG	11986
Qy	722	ATATATTAATATGATTAATGGAATTTGGAGAGAAATGCCAGAACAAATATTTGCTTATATGC	781
Db	11989	ATATATTAATATGATTAATGGAATTTGGAGAGAAATGCCAGAACAAATATTTGCTTATATGC	12046
Qy	782	CACCAAGACCCGATTTTGAATCCACTTAGGCGCAGAGGCTACATATGACAGAGATATAGGCG	841
Db	12049	CACCAAGACCCGATTTTGAATCCACTTAGGCGCAGAGGCTACATATGACAGAGATATAGGCG	12108
Qy	842	TGGAAGGTTTTTAAAGAGTATAGAGAGATGAGATTTTAAATTTGGAAGAAATCATCTAA	901
Db	12109	TGGAAGGTTTTTAAAGAGTATAGAGAGATGAGATTTTAAATTTGGAAGAAATCATCTAA	12166
Qy	902	GAGATGCGACAGCTTTGTTGGCATTTGTGATCCAAATTTCTAGATCTTTAAAAAAAAGAGC	961
Db	12169	GAGATGCGACAGCTTTGTTGGCATTTGTGATCCAAATTTCTAGATCTTTAAAAAAAAGAGC	12222
Qy	962	ACAAAGAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAAATCCTCATGAACATGAGG	1021
Db	12229	ACAAAGAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAAATCCTCATGAACATGAGG	12288
Qy	1022	CTTTAAGAAATTTATGTAAAGAAAAATGTGTTAAGTATGATGGAAGCAAAAAATGTGAAG	1081
Db	12289	CTTTAAGAAATTTATGTAAAGAAAAATGTGTTAAGTATGATGGAAGCAAAAAATGTGAAG	12346
Qy	1082	AATTCGAAATATGATTTAAACAAATCTTGCAAAATTTTCACTTCAAAAGTCATATATATC	1141
Db	12349	AATTCGAAATATGATTTAAACAAATCTTGCAAAATTTTCACTTCAAAAGTCATATATATC	12406
Qy	1142	GTCCTTTTGAATCCAAACAAAGAAATATATGAATTTGTGATGGAAGGATGTCCAACAT	1201
Db	12409	GTCCTTTTGAATCCAAACAAAGAAATATATGAATTTGTGATGGAAGGATGTCCAACAT	12466
Qy	1202	TTCTTAGCAAGAAAGATTTGTGCAATTTGAGATGCTATTTGCTATTTTGAAAAAAAT	1261


```

Db      12469 TTCTTACCAAGAAAGATTGCGAAATTAGAGTCTTATGTTCTATTTTGAATAAAAT 12528
Qy      1262 GTCCAGATGAGAAATATGCAATGTAAATATTAAGCAACATGTTAATAAAGGAACTTG 1321
Db      12529 GTCCAGATGAGAAATATGCAATGTAAATATTAAGCAACATGTTAATAAAGGAACTTG 12588
Qy      1322 ATGCACGGGCAATTAAGTGTGCAAGAAATATGCGAGAAATGTTTCAATGTTCAACA 1381
Db      12589 ATGCACGGGCAATTAAGTGTGCAAGAAATATGCGAGAAATGTTTCAATGTTCAACA 12648
Qy      1382 AAAGCTGGCTTGAAAGATTCAACAGAAATTAGTAAAGTATGTGAGAAACTGAAAAAG 1441
Db      12649 AAAGCTGGCTTGAAAGATTCAACAGAAATTAGTAAAGTATGTGAGAAACTGAAAAAG 12708
Qy      1442 AAACCAAGAGAGTTTCAACAGAAATTATTTATTTCTGTGTACAGCCAGCAAAAG 1501
Db      12709 AAACCAAGAGAGTTTCAACAGAAATTATTTATTTCTGTGTACAGCCAGCAAAAG 12768
Qy      1502 CAGCCCGTGTCTTACACATGATC 1525
Db      12769 CAGCCCGTGTCTTACACATGATC 12792

RESULT 13
AR438434 3042 bp DNA linear PAT 20-FEB-2004
LOCUS AR438434
DEFINITION Sequence 1 from patent US 6664053.
ACCESSION AR438434
VERSION AR438434.1 GI:42663289
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 3042)
AUTHORS Kovacs,J.A., Huang,S., Mauer,H., Fischer,S.H., Gill,V.J. and Mei,Q.
TITLE Identification of a region of the major surface glycoprotein (MSG)
JOURNAL Patent: US 6664053-A 1 16-DEC-2003;
FEATURES
    source             1..3042
                        /location="qualifiers"
                        /organism="unknown"
                        /mol_type="genomic DNA"

ORIGIN
Query Match 35.8%; Score 1098.6; DB 6; Length 3042;
Best Local Similarity 63.0%; Pred. No. 1.9e-178;
Matches 1954; Conservative 0; Mismatches 1054; Indels 93; Gaps 13;

2 TGCGCGGCGCGGTCAAGCGGCGAGCGAGACACAGAAATGAGCATTGATGAGAA---C 58
Db      2 TGCGCGGCGCGGTCAAGCGGCGAGGTACAGAGCATCAGAGTAGATGAGAGAGAGTGC 61
Qy      59 ATGTTTAGCTTTAATCTTAAAGAGATGACTAAGTAGACAGAAATGCAAAAAAAC 118
Db      62 GTCTTTTGGCTTTAATCTTAAAGAGATCTAAGATGATTAATAATGCGAAGAAAAAT 121
Qy      119 TAAAAAAATATGCGAAGATGACTGAGCAAAATTAATATGAAACAGATACAGAA 178
Db      122 TAGAAAAACATGCGAAGATTAAGTAGAGCAAACTTAATCCAGAAACAGTACATGAA 181
Qy      179 AACTTAAGGTTTTCGAGAGATGAGAAAGCAGATACAAATATGCAAAAGATGAAAGCA 238
Db      182 AGTTAAAGATTTCTGTGATGCAAAAAACGTGATTAATAATATGTAAGAACTTAATAA 241
Qy      239 ATATTGAGAAAAATGTAATCAATCAAGAGAAATTTAAAGAAAGCAATTAATAAAAAA 298
Db      242 ATGTTGAAAAAAATGCGGATTTTAAACAGAAATTAAGAAATGAGTGAAGAAAG 301
Qy      299 TTCAGATTTAAGCATGATGAGATGCAAGAGATGAAACAAATGCTATATTTTGGAG 358
Db      302 CTTCAAAATTTGAAAAATGATGATGTCACAAAAATGAAACAAAGATGCTGTTTGAAG 361
Qy      359 GAGTATGTTCAAAAGAACTTAAGATGATGCAATTAATTTGAGAAATTAAGTATCAAA 418

```

```

Db      362 AACATGCTC---TGATCTTACAAAGAAATTGCAACGATTTAAGAAACAAATGTTATCAGA 418
Qy      419 AGAAAGCTGATAAGTTGCGGAAGAACTTCTTTAAGAGACT---TCGAGGAGCTTTA 475
Db      419 ATAAAGCTGATAAGTTGCGGAAGAACTTCTTTAAGAGAAATTAAGAAAGAAATTTA 478
Qy      476 ATGATCATCATATGTAAGAAAAAACTTAAGAGATTTGCCCTGTCATGGGAGGAAA 535
Db      479 AAGTAAATAATTCATGTGAAAAATTAACGTGAAGTATCTCAAGAAATTAAGTCAATGA 538
Qy      536 GTGATGATTAACAACTGTGTCTGAACCAAGAGACATGTAAAGATTTTAATTTG 595
Db      539 GTGACGAATGATGAATTAATGTTTGAATCAAAAAAATAGTGTGAATTAATCTGTAAAG 598
Qy      596 AAAAAAGTAAGAGT-----CGTACTCTTTAAACAGATGTTTCAGAGCACTAG 646
Db      599 AAACGCAACAAAGTGAATCTTTCAAAAATCTTAAACGAAATTTAAACAAATTAAG 658
Qy      647 GAAGTTTAAAGAAAGAAACATGCTTGAATTTACTCGAACATGCTATTTTACATGGA 706
Db      659 AAGATGAACAACTTAATAAAAAATGCCATTTATTAATGAAGATGATTTTATGATG 718
Qy      707 ATTGCGGAGACGACGATATTAATTAATGTAATTTGAGAGAAATGCGCAAGACAA 766
Db      719 AAAGTTGTGAAACGATTAATCACTGAAGTGTAGTGAATTTGAAAAAATGTCAGAGAAA 778
Qy      767 ATATGCTTTATATGCCACCGACCCGATTTTGAATCCACTAGGCCGAGGCTACATAG 826
Db      779 ATATTAATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 838
Qy      827 CAGAGATATAGGCTGGAAGATTTTATTAAGAGTAGAGAGATGAGATTTTATTTATG 886
Db      839 CAGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 898
Qy      887 GAAAGATCATCTTAAGAGATGCAACAGCTTTGTTGATTTGATTCAGATTTCTAGTC 946
Db      899 GGAACCTTTACAGACGATGCTAGCTCTTTGGGAGATTTTGAATTCAGATCATCTC 958
Qy      947 TTTAAA---AAAAAGCAGCAAAAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAA 1003
Db      959 TTCACTCAACAGAACTTAACAAAGAAATTAATTAATTAATTAATTAATTAATTAAT 1018
Qy      1004 ATCTCATGAACATGAGGCTTTTGAAGATTTATGTAAGAAATGTTTAAGTATGATG 1063
Db      1019 ATTTAAAGACATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1078
Qy      1064 GAAAGAAATATGAGAAATGCAAAATGATTAACAAACTTGCAGAAATTTTCACTT 1123
Db      1079 AGACAAAAAATGGAAGAACTTAGTACAGATATGAAAGAAACATGTAATTTTCAATTT 1138
Qy      1124 CAAAGTCACTAATATGCTTTTTCATGATCAACAAAGAAATTAATTAATTTGAT 1183
Db      1139 CAAAAACCTTAATGATTCATTTTTCGCGATGAAATTAATTAATTAATTAATTAAT 1198
Qy      1184 GGGAGGCTTCCCAACTTTCTTAGCAAGAGATTTGCGAAATTTGAGTCTATTTGTT 1243
Db      1199 GGGGGAATTAATCACTTTCTTAAGCAATTAAGATTTTAACAAATTTGAATCGATTTGTC 1258
Qy      1244 TCTATTTGAAAAAAATGTCAGATGAGAAATGATGTAATAATTAAGCAACAT 1303
Db      1259 TTTATTTTGAAGAAAGCTGTAAG---AGCAAACTGATCAAGAAATATCAGACACAT 1315
Qy      1304 GTTAACAAAGAGACTTGAATGACAGGCAATTAAGTCTGCAAGAAATTAATGAGAGAA 1363
Db      1316 GCTACAAAGAGAGCTTGAACATTAAGCAATTAAGATTAACAAAGAAATGAGAGAA 1375
Qy      1364 TGTTAATGTTCAACAAAGAGCTGTTGAAAGTTTCAACAGAAATTAATTAAGTAT 1423
Db      1376 TGCTGATGTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1435
Qy      1424 GTGAGAACTGAAAAAGAAAGAAAGTTTCTCAACAGATTAATTAATTTCTGT 1483

```

```
Ds 1436 GCAAAAAAGTGAAGAAATGAAGATTTCCTCCAGTAAATGATTAATTTGCTTAT 1495
Oy 1484 GTGTACAGCCAGCAAAAGCAGCCCGTTGCTTACATGATCTTGGATGAAACTATCT 1543
Db 1496 GTGTACAAACATCAAAAGAGCTGATTTGCTTGGCATGATCTTGGATGAAACTATCT 1555
Oy 1544 TTTTTCGACACAACTGATCAAAAGCGAGTTTCCCGACAGATTAATAATTGCAAGGAT 1603
Db 1556 TTTTTCGACACGATTTGAAACGAAAGCAGATTTTCCAGTGAAGACATCGCGAAGAT 1615
Oy 1604 TGGGGAAGAGTCCAGATTTAGGAGAGATTCAAAAGAAATTACATGGCCATGCTA 1663
Db 1616 TATTTAAAGAAATGTAGGCTCTAAGAAAGATTTCTAATAAAATTAATGGCCATGCTA 1675
Oy 1664 CACTGAGCAGCAATGCAATCGCTTGGGGACTACAGAAATTTTAAAGCAGGTTTATGG 1723
Db 1676 CATTAAGCCAAAATTTGTGATCAATTTGAAAGCTTAAGAAATGAAAGACTTTTACTA 1735
Oy 1724 ATGAACACAAAGATCTTTGAAAGACCAAGAAAGTTGTGTAATACTTAAAGAAAGT 1783
Db 1736 ATGAACATAGAGATTAATTTGAAAAATCAAGAAATTTGGAATGTATTTGAAAGGAAT 1795
Oy 1784 GTAATAATGGCTGAAGAGGAGATGACCGTTTCTTTTGTATGTGCTTCCAAAAG 1843
Db 1796 GCAATGAATGGCTGAAGAGGAAATGAACGTTTCTCTTTTATGTGCTTTCGAAATA 1855
Oy 1844 CTACGTGTAGCTGATGTGTAAGAAAGCTGAAGACAGGTGTGAATTTCAAAAATA 1903
Db 1856 GGACTTGCAAGATATATGTGAGAAAGATGTGAATAATCAATGCAAAATATTTGAAAAACA 1915
Oy 1904 TAAAGCTTCATATATTTATGAATTTCTTGAATAATTAACAATTAATAACACACTGG 1963
Db 1916 TTAATAAAATACCAAGTATG-----ATAGTAAACTAAATAGAAAGATAG 1963
Oy 1964 AAAGAAATGTCCTCTTGCCATACGTAATGCTAATGCTTCACTAATTTGTCAGGTC 2023
Db 1964 GGCATATTTGCTTATTTGGACCCACACTGCCATAGATTTGGACCCATTTGCCGGAATC 2023
Oy 2024 TTACGAAAGAGATGTTGTACAAAAATCAAGAACATCGTAGCCGTTCTATTAAGAA 2083
Db 2024 TTGAAGAAA---AATAAATGTGAAGCTTTGAAAAATATTTGCAAACTTATTAAGCAA 2080
Oy 2084 AGGCTTGGAAGATGCTCTCAAGTAGAGCTTCAAGAAATTTGACTGATTAATTTAAAT 2143
Db 2081 GAGACCTTGAAGAAATCAGCTTATATTTGATTAGTGAGCATCTTGATTAAGAAAAAACT 2140
Oy 2144 GTGAACCTGCATTTGAAAGATATTGTACAGTAGCGGGAAGCTAATAATATGTCATCA 2203
Db 2141 GCAAAACAAATCTTGAATAGTACTGTACCTATGGAATCAACAGAAATTAACACATTA 2200
Oy 2204 GTGGCTTATGCAAAGCTTAACCAAGATTAACCTTGAAAGAGTAGAGATGCTAGAA 2263
Db 2201 AAGTTTTTTGTA-----ACAGTTTCACTGATTAACAATGAACATTAGAG 2245
Oy 2264 AGGAACCTCTGAGAAATTTAGTGAAGAGTGAAGAACAGTGCACAAAGCTTACCAACAG 2323
Db 2246 ATAAACTTTGCAAAACCTAGTTCAAGCGTGTAAAGAAAAAGCCAGAGATTAATCAAAAG 2305
Oy 2324 AATTAGGACAACCGGAGCTGATCTAATAAAAGATTATTAAGCATATGAGGAAGCTTAAG 2383
Db 2306 AACTTGAAGAAAGCAAAATATTTAGAGAAACATTAAGATTAAGAAAGTAAAGAA 2365
Oy 2384 AACGTGACAGAGAACATGAACAGTCCAGTCTTGTGTTGCTCACTTAAGAAAAAG 2443
Db 2366 AGGATATCAAAAAATTCATGAGAAACAAATCTGTTTTTCAACAATAATCAACG 2425
Oy 2444 AAAGTATGTATCAAAAAAGTAATGCAAAAAACAGATTAAGATGCGCTTCAACGAGC 2503
Db 2426 ATAAATAAACAGAAAAAG-----AGTCAAGCTATGATCGCTTACTGTAG 2470
Oy 2504 TTCAAGATACCAAAACATGTGAATACTACGAGAGGGAAGTTAAGATGATACCGTAA 2563
Db 2471 TTCAAGATATTTGTACATTT-----AACTGTAAAGAAATGAAGAAAGTTCAAGTGA 2524
```

```
Oy 2564 CAGATTAGAAAGCTAAAGCTTTGATTTGGCAGAGAAATATTTGAAAGATATGTAGATT 2623
Db 2525 CAGAAAAAGAAAGCAAAAGCGTTGATTTGGTACATGACATTCAGTCTTATGTAGAGT 2584
Oy 2624 TGAAGAAAGATGTAAATTAATTTGGAATCAATTCGACAAATTAAGAGAGATTGCAAAAGCT 2683
Db 2585 TAAAGAAACGTGTACCATCTTAAGAGATGATTCGAATTTTGAAGAAAGATGTAAA--T 2641
Oy 2684 TAGAAGAAATGTGCAAAAGATTAATTAAGGCTTTCGCAATCTGAAGCCTGTAGAGTGA 2743
Db 2642 GTAAAGACCAATGCAAAAGATTAAGAAATATGTTTAAATAATGAACCACTGAAAGTAA 2701
Oy 2744 AGCCGCAAGAAACGTGACAGAAAGTACCAACGACACTACACAAACAAACACCGTTG 2803
Db 2702 AGCCACATGAATTAATAAACAGTAAACGAAACCAATACAAACAGTCCAGAAAAAGTCA 2761
Oy 2804 CCGATCCGAAAGCAACG-----AATGCAATTCCTTACAGCAACAGCA 2848
Db 2762 AAGAAGCAAGAAAAACAGTAGAGACGAGAGAAATGCAAAATCTCAGCACAAACAGCA 2821
Oy 2849 CATGGGTTTACACAGACATCGACACACAAAGCAGTCTACTATGACATCTACATCACAT 2908
Db 2822 CGTGGTTCACAAAGACGTCAACCCATACAGACCTCCAGACATACCTCCACAGTTACGT 2881
Oy 2909 CAAAATTAACATTGACATCAACGAGCGATGCAAAACCAACAGTGTACGACAGAGAGG 2968
Db 2882 CAAGATTAACACTGACCTCGACGAGCGGTGTAAAGCTTACGAAGTGTACGACAGAGAGG 2941
Oy 2969 AAGATGATGACGAGACAGTGAACCGAGTGAAGGGGCTGAGAGATGACGCGGTGAATGGA 3028
Db 2942 AAGATGAAGCAGAGAGGTGAAGCGAGTGAAGGGGCTGAGAGATGAGTGGTGAAGTGA 3001
Oy 3029 TGAGGGGGGTGATAGTAGCAATGTTATTTGTTCAATGATT 3069
Db 3002 TGAGAGGGGTGTTATTAGCAATGATGATTTTCAATTCATGATT 3042

RESULT 14
AR438435
LOCUS AR438435 3006 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 3 from patent US 6664053.
ACCESSION AR438435
VERSION AR438435.1 GI:42663290
KEYWORDS
SOURCE
ORGANISM
Unknown.
REFERENCE
1 (bases 1 to 3006)
AUTHORS Kovacs,J.A., Huang,S., Maabur,H., Fischer,S.H., Gyll,V.J. and Wei,Q.
TITLE Identification of a region of the major surface glycoprotein (MSG)
JOURNAL gene of human Pneumocystis carinii
Patent: US 6664053-A 3 16-DEC-2003;
FEATURES
location/Qualifiers
1..3006
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 30.2%; Score 927.4; DB 6; Length 3006;
Beet Local Similarity 60.3%; Pred. No. 4,1e-149;
Matches 1859; Conservative 0; Mismatches 1131; Indels 93; Gaps 16;

Oy 2 TGGCGCGGCGGTCAAGGCGAGGCGAGGACAGACAGATAGCATTTGATGAGAACTAG 61
Db 2 TGGCGCGGCGGTCAAGGCGGCGGCTGCAG---CACGAAATAGTGTGAAAGAAATATC 58
Oy 62 TTTTACCTTAATTTCTAAAGAGATGACCTAAGTAGACAGAAATGCAAAAAAACTAA 121
Db 59 TTTTGGCTTTGATTTTGAATAATGATGAATAATTAATGATTAATGTAATAAAAGTTGA 118
Oy 122 AAAAAATTTGCCAAGATTTGACTGAGCAAAACCTAAATATAGAACAGTACACAGAAAC 181
```

Db 119 AAGAGTATGTGAGTTTAAAAAATGTACAAAGAAACCAAAAACTAGAAAGAAAGT 178
Qy 182 TTTAAGTTTTTGGGAAGATGGAAGC---AGATCAAAATTCGAAGAACTGAAGCCA 238
Db 179 TAGCGGAATCTGCAAGATGATATAACATATGAAGCAAAATGCAAGAAATCAGAAACAA 238
Qy 239 ATATTGAGAAAAATGTACTACAAATCAAAAGAAAACTTAAAGAACTTTAAAAAATA 298
Db 239 AGGTTAAGCAAAAGTACTAGTTTTCBAACGAACCTTGATTAAGCACTCAAAAAGGAG 298
Qy 299 TTCAGATTATACGGATTAAGGATTGCAAGAAATGAACCAATGCCATTTTGGAGG 358
Db 299 CTTCAACATTATGAAGATATGATTGTAAAGAAATGAACGACAACTCGTGTTTTGGAGG 358
Qy 359 GAGTATTTTCAAAAGAACTTAAAGATGATTGCCAATCTTTAGAAATTAAGCTATCAAA 418
Db 359 GAGATGTCCCAACGAACTTAAAGATTAATGTATGAACGAGAAATTAAGTTATCAAA 418
Qy 419 AGAAACGTATTAAGTTGCGGAAGAACTTTTAAAGCACTTCGTAGCGATCTTAATG 478
Db 419 AAAAACGAGACGAGTAGACAGAAAAAGCTCTTTAAGGTACTTAAGAGGAACCTTAAGG 478
Qy 479 GATCAGTCATATGTGAAGAAAAAACCCTTAAGAGATTTGCCCTGATGGGAGGAAAGTG 538
Db 479 ATAAAAACACATGCAAAAAATTAAGTTAAAGGGGTTGTCAAGAAATTCAAAGAAAGTG 538
Qy 539 ATGAGTTAACAACTTGTGTGTAACCGAAGAAAGACATGTAAATTTTAAATTGAAA 598
Db 539 ATGAGCTAAATTAATATGTCTTGACGAAGAAAAACGTGTGAGATCTTGATCTAAGA 598
Qy 599 AAGATTAAGAGTGGGACTCTTAAACAGATGTTTCAGCAGCACTAGAA-----GTT 652
Db 599 AAGATTAACAAATGCAAACTCTCAAGAAAGAAATTTGATCTAGCTTGGAAGAAAGATTT 658
Qy 653 TTTAAAAAGAAACATGCTTGAATTAACGAAATGCTATTTTACATTTGAAATTCG 712
Db 659 TATTTAAAAAGAAATGTTTATTTATTTCTGAAGATGTTACTTTTATGGCTCAAACTGTG 718
Qy 713 GAGACGAGATATTAATTAATGTATGAAATTGGAGGAAAAATGCCAGAAACAAATATTTG 772
Db 719 AAGACAGATCAGCCA---AAGGTAAAGAGTTTGCAAGCAATGTCAAAAGGAAATCTCG 775
Qy 773 CTTATATGCCACGAGACCCGATTTTGAATCCAACTAGGCCAGAGGCTCAATAGCAGAG 832
Db 776 TTTATGCAACGACCGAGTTCAACCTTGATCTTAAGAAATTAAGATTAAGTTAGTACAGAG 835
Qy 833 ATATAGGCGCTGGAAGATTTTATTAAGAGGTAGAAGAGATGAGTTTATTTTGAAGA 892
Db 836 AATATAGACCTAGAAAAATTTGATCGTAAGAGAGTGAAGAAAGGAATTCATATTGGAAGG 895
Qy 893 ATCATCTAAGAGATGCCACAGCTTTGTTGGCATTTGTCACAAAGATTTCTTAATA 952
Db 896 CATCAATTAAGATGAAGTCCGTTATTTGGCATTAATTAGCA-----GAGTATGCTC 949
Qy 953 AAAAAGACGAAGAAAGAAATGCGAAGAGCCCTTCAAAAAAGCGCAAAATCTCATG 1012
Db 950 AAAATTACTTTTAAAGATCAATGTGAAGATGTTTATTAAGAAATTAAGTGAAGAACTTTAAG 1009
Qy 1013 AACATGAGGCTTTAAGAAATTTATGTAGAAGAAATGTTTAAAGTATGATGGAACGA 1072
Db 1010 AGCATATTTATTTTAAAGATTTATGTAGTATTAAGATCACTGATTAATCCAAAAAGAAA 1069
Qy 1073 AATGTGAAGATTTGCAAAATGATATTAACAAACTTGGCAAAATTTTCACTTCAAAAGTCA 1132
Db 1070 AATCGAAGAACTAAATTAAGAGATT-----ACCAACCCGTAATTTTAACTGTTTCTAAA 1123
Qy 1133 CTAATTAATCGCTTTTGAATCAACAAAGAAATTAAGAAATTTGATGATGGGAAGGCT 1192
Db 1124 GGAATGGAATAATTTCCCTCAGCTAATGTAAA---GAAATTAATTTGTTGGCATATGT 1180
Qy 1193 TGCCAACTTTCTTAGCAACGAAGATGTGCGAAATTTGAGTCTTATTTGTTATTTTG 1252
Db 1181 TGCAATACATTTCTTGTGTAAGAGAGTGTACGAACCTGTTGTGGATTTGTTTATTTGA 1240

Qy 1253 AAAAAAATGTCCAGATGAGAAAAATGATGTAAAAATATTAAGCAACATGTTACAAA 1312
Db 1241 AAGCCAGCTCCACTT---GAAAGCCCTGCAATTACTTAAAGACGATGTTATATAAA 1297
Qy 1313 GAGACTTGTATGCAACGGGCAAAATTAAGTGTGCAAGAAAAATATGCGAAGATTTACATG 1372
Db 1298 AAGGCTTGAAGCAATGCAAAATGAAGCATTAAGATTAAGTTACGGGAAAAATTTGCAAG 1357
Qy 1373 GTTCAACAAAGCTGCTTGAAGAAATTTCAACAGAAATTAAGTAAAGTATGTGAGAAAC 1432
Db 1358 GTTCAAAATTAACATGCTTGAAGAACTTCAAAAAAATTGTGAAGAAATTTGTGAAAAAGA 1417
Qy 1433 TGAAGAAAAAGAAACAAAGAAATTTCTCAACAGATGAATATTATTTCTGNGTGTACAG 1492
Db 1418 CGAAGAGCA-----AAGTATGATTAATTTTGTACTATGTATGAACC 1459
Qy 1493 CAGCAAAAGCAGCCCGGTTGCTTACACATGATCTTCAATGAAGAAATCTATCTTTTACGAC 1552
Db 1460 CAATTAAGAAACGGCTCTTACAGTGTCAACAGATTTGCAATGAAGGAGCAGTTGCTTGCAG 1519
Qy 1553 AACACTGATCAAAAGCGAGATTTCCCGACAGATTAATAATTGCAAGAAATTTGGGAGAA 1612
Db 1520 AGCATTTGAACGAAAAAGAGATTTTCCAAAGAAAAAGATGTGAAGAAATTAAGAGAAA 1579
Qy 1613 AGTCCCAAGATTTTAGAGAGAGATTCAAAAGAAATTAATCATGSCCATGTATACACTGGAGC 1672
Db 1580 AATGTAGGCTTTTAGAGAAAGATTCAGAGAAATTAATAGTCAATGTATTAACGTTAAAC 1639
Qy 1673 AGCAATCAATTCGCTTGGGACTACAGAAATTTTAAAGCAGTTTATTTGATGAACACA 1732
Db 1640 AGCATTCATTCGGCTGAAAGACATAGACACTTAGAAGAGAGTTGCTTAAAGAAATA 1699
Qy 1733 AAGTACTTTGAAGAACCAAGAAATTTGTAAATCTTAAGAAAGAAATGTATTAAT 1792
Db 1700 AAGGATTTTAAAGATGAAGAAATAGCTCAAGAGAGCTTAAGAAACGATGTGAAGAAAT 1759
Qy 1793 GGTCTAGAGAGGAGATGACCGTTCTCTTTGATGTGTCTTCAAAACGCTACGTTG 1852
Db 1760 GGTTAGAGAGAAATTAATTAATTTTTCGCTTGTCTGACTTGGAACTTTTGA 1819
Qy 1853 AGCTGATGTATTAAGACGTAAGAAAGACAGTGTGAAGTATCAAAAAAATATTAAGCTT 1912
Db 1820 AAAAGATCACTAGAAATGTTGAATCTAATGTATATATTGAAGAGACATATGGAACATA 1879
Qy 1913 CATATATTATGAATTTCTTGAAGATTAACAAATTAATTAATTAACACACTGGAAGAAATT 1972
Db 1880 TGAACGTTATAGTAAGTAATGCTTAAGAAAGGAAAAAATATGTGAATT-----TT 1930
Qy 1973 GTCCCTTGGCATATGCTATTTGCAATGATTTTCACTTAATGTCCAGGCTTACGAAG 2032
Db 1931 GGGCTCATATTTGTAAAGAAATGACAGCAAAATTTGTGAAGAACTTAAACCGAGGAAG 1990
Qy 2033 AGAATAGTTGTACAAATCAAGAAAGCATGTGAGCCGTTCTATTAAGAAAGGCTTTGG 2092
Db 1991 ATGGGCAATCGAAAAATCTCAATTAAGAAAGTCAAAATCATTTCTTGAAGAAAGGCTTTAG 2050
Qy 2093 AAGATGCTTCAAGATGAGCTTCAAGAAATTTGACTGATTAATCTTAATGTGAACCTG 2152
Db 2051 AAAATTAAGTTGTGAAGAAATTTGAAGTATTTATCAAAAGTATGAAGAAATGTAAACATA 2110
Qy 2153 CATTGAAGAAATTTGTACAGTAGCGGGAACGTAATTAATGCTCAATGAGTGCCTTAT 2212
Db 2111 CACTTAATATATCTGTACCAATTAAGAAAGGAGAAATGGGTGGAACCTTTGT--- 2167
Qy 2213 GCAAGCTTAACCAAGAGATTAACCTTGAAGAAAGTATGAAGTCTAGAGAAAGAACTCT 2272
Db 2168 -----GCAAGAGAAAGAAAGAAC---CAAGGTGACATTTAAAGTTAAGAGAAAGAACTCT 2218
Qy 2273 GTGAGAAATTTAGTGAAGAAAGTGAAGACAAGTGAAGCAATTAACAAAGAAATTTAGAC 2332
Db 2219 GTGAAGAGCTTAATTAAGATTAAGAAAGAAAGAAATGCTCAAAATTTGAAGAGAGAGCTTGAAG 2278

QY 2333 AACGGGACGTGATCTATAAAGATTTATAGACATATGAGAACTTAAAGAACTGCG 2392
DB 2279 AAGTAAAGAGGTCTTAAAGAAAGAAAGAAAGTATTAATAAAGAAAGAAAGCAG 2338
QY 2393 AGGAAGCAATGAAACAAAGTCCAGCTTGTGTTGTCTCATCTATTAAAGAAACGAAGTAATG 2452
DB 2339 AAAAAGCCATGGAAGATGCAACCTTATTTATTCAGAGAGCGAAAGACCTGATTAAT- 2397
QY 2453 TATCAAAAAGTAAATGACAAAACAAAGATTAAGATGCCGTTTCAACCGACTTCAGATA 2512
DB 2398 -----AATTAATTAAGTCAAGTAAATAAAGACTCATCTGATATACCTTAAGAGAAAGAAAGCA 2452
QY 2513 CCACAAAACATGTGAAATTACTACGAGAGGAGTTAAGAGTATCCCTAACAGAAATAG 2572
DB 2453 CAACGAGATTT-----AACTTGTAAAGAAAGAAAGTGAATGATTAACAGAAAG 2506
QY 2573 AAGCTAAAGCAATTTGATTTGGCAGACAGAAATTTTGAAGATATGTAGATTGAAGAA 2632
DB 2507 AATTAGCAGCATTTTATTTGTTGGTAGAGAGCATTTGATCTTATCTAGATTGAAGAA 2566
QY 2633 GATGTAATTAATTTGGAATCAATTCAGATTGCAAAATTAAGAGGATTTGAAAGACTTAAGAAAG 2692
DB 2567 TATGTAATCATTCATCTAGAAATTTGGTTTCAAAAAGAGTG--TGACTGTGAGATC 2623
QY 2633 TATGCAAAAAGATTAATTAAGCTTGTCCGCAATCTGAAGCTCTGAGAGTGAAGCCGACG 2752
DB 2624 CATGTAAAAAAGATACAGGAAATATGTTCAACATTAGACCACTAAATAGACACACACG 2683
QY 2753 AAACAGTACAGAAAGTACCAACGACCACTACACAAACA-----ACAAACACCGTTGCG 2806
DB 2684 AATATGTAATTAATAAACAATTAACATCAACCAACCAACCAACCACTACATTAAG 2743
QY 2807 ATCCGAGGCAACGGAATGCAAAATCTTACAGACAAACAGACATGGGTTACACAGACAT 2866
DB 2744 ACGCAAGGCAACAGACTGCCACTTTACAGACAAACAGATACGTTGGTCAAAAGCGT 2803
QY 2867 CGACACACACAGACGCTCTACTATCAATCACTACATCAATCAATCAATCAATCAATCAAT 2926
DB 2804 CGACCCATATCTAGACATCTCAACACCACTACAGTCAACGTAACGTTGACCT 2863
QY 2927 CAACGAGGCGATGCAACCAACCAAGTGAACGACGAGAGAGAGATGATCAGAGACG 2986
DB 2864 CGACAAACGCTGTAAAGCTTACGAAGTGTACACAGAGAGAGATGAACGAGACG 2923
QY 2987 TGAACCGAGTGAAGGCTGAGAGATGAGTGGTGAATGTGATGAGGCGGCTGATAGTAG 3046
DB 2924 TGAACCGAGTGAAGGCTTGAAGATGAGTGAAGTGAAGTGAAGGCGGCTTATTAG 2983
QY 3047 CAATGTTATTTCTTCATGATT 3069
DB 2984 CAATGACGATTTTCATTCATGATT 3006

RESULT 15
AF149015 3459 bp DNA linear PLN 20-JUL-2000
LOCUS AF149015 3459 bp DNA linear PLN 20-JUL-2000
DEFINITION Pneumocystis carinii f. sp. ratti major surface glycoprotein R7a
(R7a) gene, partial cds; and major surface glycoprotein R7b
gene, partial cds.
ACCESSION AF149015
VERSION AF149015.1 GI:9294725
KEYWORDS Pneumocystis carinii f. sp. ratti
SOURCE Pneumocystis carinii f. sp. ratti
ORGANISM Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 3459)
AUTHORS Schaffzin,J.K. and Stringer,J.R.
TITLE The major surface glycoprotein expression sites of two special
forms of rat Pneumocystis carinii differ in structure
JOURNAL J. Infect. Dis. 181 (5), 1729-1739 (2000)
MEDLINE 20283743
PUBMED 10823775

REFERENCE 2 (bases 1 to 3459)
AUTHORS Schaffzin,J.K., Garbe,T.R. and Stringer,J.R.
TITLE Direct Submision
JOURNAL Submitted (06-MAY-1999) Molecular Genetics, Biochemistry, &
Microbiology, University of Cincinnati College of Medicine, 231
Bethesda Ave, Cincinnati, OH 45267-0524, USA
FEATURES
source
1..3459
/organism="Pneumocystis carinii f. sp. ratti"
/mol_type="genomic DNA"
/db_xref="taxon:38082"
/note="forma_specialis: ratti"
1..500
/gene="R7a"
1..500
/gene="R7a"
/product="major surface glycoprotein R7a"
1..500
/gene="R7a"
/codon_start=3
/product="major surface glycoprotein R7a"
/protein_id="AA06638.1"
/db_xref="GI:9294726"
/translation="TTTETETVTTTETFEVLSPENSTVPGKTIENITTIKMTNTEIK
YSSPIITTKFTIENCTTICDINTQFTKTSFQITVTSVTTTIIISITKLTSLK
CTPKSTSDSTKETQPEPTKSTSDSTKETQKEEVAKNEGVKIRISEMTKIFFGIT
IMGL"
629..903
/note="similar to the upstream conserved sequence (UCS)
exon II"
1..904
/gene="R7b"
1..904
/gene="R7b"
/product="major surface glycoprotein R7b"
1..904
/gene="R7b"
/note="silent donor MSG that in order to be expressed must
be linked to a unique expression locus (UCS) whose coding
sequence possesses an ATG start codon"
/codon_start=1
/product="major surface glycoprotein R7b"
/protein_id="AA06639.1"
/db_xref="GI:9294727"
/translation="AAQVPRERAPKADPKEPAVPGPSPFGIQQVVAEGIGEE
LILALIVSEYNNENCKTALKKYCDGLNEDPVLQVNEKVKGLCKDKEKCKDL
KGEVKEKVEFGKQLEKALKTGDEKCKLEBKCIPLSEADPNLKENCYLNECY
KBRKDVABEVLIRGLKVGSGSVKDCGKLEBVCISLGSYDELMPFLCDNNCTNL
NTRVNYCTPLKEIKVNSKEBEKLEDCYFVSSCGDAKKCKEIKKQCEKFTYT
APSGSPVERKRGFEKIGLEEVLRABADGVLVGRPGTIRMLKDLILULSKGTGV
QKTCENVLKKNCDFRYLQELKGLCGDNGRKCKBELVKDVSSRCSYLKYLPLK
GLSTTPDGKTSNVLIMGELPTLLSNEDCALESCELVESCKNNLEKACGNVAAQY
KKGQDVRVNLAFEGRLBAHNSSLSGGQVCEVIVIEKCEKLEKREYVAKCLFPGELC
QQVLEDFIVRLBELGRVLDVDRPREDDCVLELKKCSHLGNVRLNEDKCATPETYC
HYLAVSQLEKDPKRSVLIISSEKKEGKERSCDLNRKRKPSVSLAPETCY
PFSNEIRIECIFIQRVEDLKNNGQDOBOVCAPFAYIDYGVCSITLCKNGKDN
GPFPELKEGPFPELLEBELLYELGSLNISECKKLEKCTQFKNSTNDLTGL
CKDNGNENBEVCKLKVKEBKCPSLTNTLEKAAVLEBKKEVDEYKAEKRPNTA
SLVSRPKGKEPAPAEKPAPEKPAAGSAAANGITVKADVDASVYRIPTPR"

ORIGIN
Query Match 13.0%; Score 399.8; DB 8; Length 3459;
Best Local Similarity 51.6%; Pred. No. 1e-58;
Matches 1242; Conservative 0; Mismatches 1062; Indels 105; Gaps 10;
QY 44 GCATTGATGAGGACATGTTTACCTTAATTTCAAGAAAGATGAGCTAGACGACG 103
DB 1022 GCATTGGGAGGAACTTATTTTGCCTTGTGTAAGAGTGAATTAATTAATGAAGCA 1081
QY 104 AATGCAAAAACCACTAAATATTTGCCAAGATTTGACTGAAGCAAAAATAATATAG 163
DB 1082 ATTGCAAAAACGCCCTCAAGAAATATTGTGATGATGAATGAACAGATCGGAACCTTAC 1141

QY 164 AACAGTACACAGAAAATTAAAGTTTTCGAGATGAGAAAAGCAGATCAAAATGCA 223
 DB 1142 AGAAAGTGAATGAAAAGTTAAAGGACTTTGTAAGATTAAGAACAGAAAATTAATGCA 1201
 QY 224 AAGAACTGAAAGCCCAATTGTGAGAAAAATGTACTACATCAAAAGAACTTTAAAGAG 283
 DB 1302 AAGCTTGAAAGAGAACTTGAATAAAAAAGTTGAGAGTTTAAACAGAACTTCAAGGTG 1261
 QY 284 CAATTAATAAAAAATTGAGATTATTAACGGAATTAAGATTGCAAAAGAAATGACAAAT 343
 DB 1262 CATTGAAAAAACAAGCGA-----TGAGAAAGATTGTAAAGAACTTAAAGAAAT 1312
 QY 344 GCCATTATTTTGAGAGAGTATGTTCAAAAGAACTTAAGATGATTCGAATCTTTGAGA 403
 DB 1313 GTATATTTTGAAGAGCGGATCCAACTAATCTTAAAGAGAACTGGCTAAAGTTGAGGA 1372
 QY 404 ATAACTGCTATCAAAAGAAAGTGTATAAGTTGCGGAAGAGTTCTTTAAGACATTC 463
 DB 1373 ATGAATGCTATTAAGAAAGCGAAAGATGTGCGAGAGAGTTCTTATTAAGAGTCTTA 1432
 QY 464 GTAGCGATCTTAATGAGATCAGTATGTGAAAAAAAATTAAAGAGATTGGCCCTGCA 523
 DB 1433 AAGTTGTTGGCGGTAGTGTAAAGATTGCAATGGAAAAATAGAGAGTTGTGCAAAAT 1492
 QY 524 TGGGGAAGGAAAGTATGATGTTAACTTAATGTGTCTGACAGAAAGACATGTAGA 583
 DB 1493 TAACTGGATACAGGATGAATTTGATGTGTCTTGAATCCAAATTAATGCGTCAAA 1552
 QY 584 ATATTTTAAATTGAAAAAGATTAAGAAAGTGGTACTCTTAAACAGATGTTTACAGACAC 643
 DB 1553 ATCT-----CAATACAGAGTGAACAAATGTTTGCAACCCCTTTAAAGAAAGAAC 1600
 QY 644 TAGGAGTTTAAAAAAAGAAACATGTCTGAAATTAATCTGAACTATGCTATTTTACATG 703
 DB 1601 TGAATAATGTAATCAAGAGAAATGTACGAAAGCTTGAAGACTGCTATTTTACGAT 1660
 QY 704 GAAATTCGGAAGACGACATATAATTAATGTATGTAATTTGGAGAAATATGCCAAGAAC 763
 DB 1661 CAAGTTGTG-----CGATGCAAAAAATGTAAGAAATTTGAAAGCAATGCCAGGAAA 1714
 QY 764 AAAATATTGCTTATATGACCAAGAGACCCGATTTTGAATCCAACTAAGCCAGAGCTACAA 823
 DB 1715 AAGATTTTACATACACCGCCGAGATCGGTTTTCAGTCCGCTGGAAGAAAGAACAGAGAT 1774
 QY 824 TAGCAGAGATATATGGGCTGGAAGGTTTATTAAGAACTAGAGAGAGATGAGATTTTAA 883
 DB 1775 TCGAGAAAAAGATTTGGTTGGAAGGTTCTTGAAGAGCAGAAAGCAGATGGGTTTGG 1834
 QY 884 TTGGAAGAAATCATCTAAGAGATGCGACAGCTTTGTCGATTTGATCCAGATTTCTA 943
 DB 1835 TTGGAAGACCGGATACAAAGATGCTGAAGAAATTTGATTTCTGTTGTGACAAAGGGGAGA 1894
 QY 944 GTCTTAAAAAAAAGACGAAAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAA 1003
 DB 1895 CTGCTAACCTTCA-----AAAAACAATAAGCGAAAAATGTGTTAAAAAAAATTTGAT 1947
 QY 1004 ATCTCATGAACATGAGGCTTTAGAAAAGTTTATGTAAGAAAAATGCTTAACTATGATG 1063
 DB 1948 AAATTTCAAGTCTGAATCAGAG-----TTAAAGGGGTATGCGGTATCAATATGTA 2002
 QY 1064 GAAAGAAAAAATGTAAGAAATTTGCAAAATGATATTAACAAACTTGCAAAATTTTCACTT 1123
 DB 2003 GACAAABAAAATGTGAAGATTTAGTAAAGATGTCTATCCAGATGTTCAATCTTAAT 2062
 QY 1124 CAAAAGTCACTAATATATGCTTTTGTATCAACAAAGAAATTAATGAAA--TTGTTG 1180
 DB 2063 ATAAAGCTTTTGTGAAGGGGTGTCTACGACATTTGATGTAACAAACGTAACCTTCTGT 2122
 QY 1181 GATGGGAAGGTTGCAACATTTCTTACAGAAAGATTTGCGCAAAATTTGAGTCTATTT 1240
 DB 2123 TATGGGAGACTTCCAACTTGTCTAGCAATGAAGATTTGCGAGAGCTTAAGTCTGAT 2182
 QY 1241 GTTCTATTTTGAATAAAAAATGTCA--GATGAGAAATGATGTAAAAATATTAAGAG 1297

DB 2183 GTTATACTTAGAGAGTTCAATGTAAAAATATCTAGAAAAAGCATGCAAAAATGTGAGG 2242
 QY 1298 CAACATGTTCAAAAGAGGCTTGATGACAGGCGAAATTAAGCTCCAGAAATATGTC 1357
 DB 2243 CAGCTGTATTAATAAGGAGCAAGACAGGCTGTGAATGCAATGTTTGAAGAAAGTTAG 2302
 QY 1358 GAGGAATGTTACATGTGTTCAAAACAAAAGCTGCTTGAAGTTTCAACAAATTAAGTAA 1417
 DB 2303 AAGCTCATGGAAT-----TGAGTTGCTTGGATCAAGGCG 2338
 QY 1418 AAGTATGTGAACCTGAATAAAAAAGAAACAAAGAGTTTCTCAACAGATGAATTAATTA 1477
 DB 2339 AAGTGTGAAGTATGATGATGAATAAATGTAAAGAACTTAAGATTAACAGAAATACCTG 2398
 QY 1478 TTCTGTGTATACAGCCAGCAAAAGCAGCCGCTTCAACATGATCTTCAATGAGAAA 1537
 DB 2399 CAAAATGCTTTTCCGGAGAAATTAATGTCAACAAAGTTGAGAGATATTTTGTAGAT 2458
 QY 1538 CTATCTTTTACGACAACTGATCAAAAAGCAGATTTTCCGACAGATTAATAATGCA 1597
 DB 2459 TAGAGAAATTAAGGAGAGTTTGTGATGATGAGAGATTTTCCAAAGAGAGATGATGCG 2518
 QY 1598 AGGAATTTGGGAGAAAGTCCAGATTTAGAGAGATTAACAAAGAAATTAATGATGCCAT 1657
 DB 2519 TTGAATTTGGAAAGAAATGACAGCCATCTGGAATTAATTTGCGCTTGAATGAATGAAT 2578
 QY 1658 GTCATACATGAGACCAATGCAATGCTTTGGGAGATCAACAAATTTTAAAGAGGTTT 1717
 DB 2579 GTGCTAATTAAGTGAAGCGCTGCTATTAATTAATGCTTTTCAAGACAAATGAGGAAAGTT 2638
 QY 1718 TATGATGTAACAAAGATTAATTTGAAGACCAAGAAAGTTGTGTAAATTAATCTTAAAG 1777
 DB 2639 TCTTAAAAAGAAACAGATGATTAATTAAGTAGAAAAATGTAAGAAAGTTTGAAG 2698
 QY 1778 AAAAGTGAATTAATGCTTGAAGAGAGATGACGTTTCTCTTTGTATGTGCTTCC 1837
 DB 2699 AAAAGTGTATCAACTGAATTAAGAGAGAAAGAACCCGTTTATGTTGCGCCATATGC 2758
 QY 1838 AAAACCTACGCTGACCTGATGTTAAAGACGTAAGAACAGCTGTGAATTAACAAA 1897
 DB 2759 CAGAAAGAAACATGTAATTTTCTCAATGAATAAAGATTAAGTGTCTTATATTTAAAG 2818
 QY 1898 AAAATATTAAGCTTCATATATTAATTTGAATTTCTTGAATAATTAACAAATTAACAA 1957
 DB 2819 ATAGGTTGAAGATGATGATTAATAATTAAGCAAAAGATGACAAAGTTTGCACCGCT 2878
 QY 1958 CACTGAAAGAAATTTGCTCTTGGCATAGTATGCAATGATTTTCACTTAATGTC 2017
 DB 2879 TCGCTCATATTAATGATGAATATGAAGAAAGTTTGTCAAGAAATCTT-----AGGAA 2929
 QY 2018 CAGCTTTAGCAAGAAAGATATGTTGTACAAAAATCAAGAAAGATCGTGAAGCCGCTTATA 2077
 DB 2930 AAGGTAATGTGAAGATTAAGTGAACCTTTTAAGAACTTGAAGAAATGTGTGCAATCTTGG 2989
 QY 2078 AAAGAAAGGCTTGAAGATGCTCTCAAGTGAAGCTTCAAGAAATTAATGCTATTAAT 2137
 DB 2990 AAGAAATTAATTTGAAGAGAGATGATTAATGAACCTTAAGAGAGATTTAACCAATGATA 3049
 QY 2138 CTAAATGTGAACCTGATTAAGAAAGATTAATGTAACATGAGGGAAGAGTAATTAATGCT 2197
 DB 3050 GTGAATGTGAAAAAATTTGAAGATTAATGTACTCAATTAACAAATTTCAACAAATGATA 3109
 QY 2198 CAATCAGTGTGTTTGAAGAGCTTAACCAAGATTAACCTGGAAGAGATGATGAGATG 2257
 DB 3110 CATTGACTGTGTTATATGAAGATTAAG-----ATGGCG 3142
 QY 2258 CTAGAAAGAACTGTGAGAAATTAATGTAAGAAAGTGAAGAACAGTGCAAAGCATTAAC 2317
 DB 3143 GGAATGAAGAGTTTGCAAAATTAATGTTAAAAAGTGAAGAAATGTCTCTTTAA 3202
 QY 2318 CAACGAATTAAGCAACCGGACAGCTGATTAATAAAGATTAATTAAGCATATGAGAAC 2377

Db	3203	AAACTAATCTGGAAGCGAAAGCTGAATTGGAAGAGAAAAAGAGGAATACGATGAGG	3262
Qy	2378	TTAAGAAACGTGACAGAGGAGCAATGAAACAAGTCCAGTCTTGTGTTGTCACTCATTAAGA	2437
Db	3263	TGAAAAAAGAGCGCAAAAAATTGCGGAATACAGCTAGCTGTGTGTGCGAGACCCAGAGA	3322
Qy	2438	AAAACGAAA	2446
Db	3323	AAGGTGAGA	3331

Search completed: October 29, 2005, 06:20:32
Job time : 12899 secs

QY 1861 GTAAAGACGTGAAGACAGGCTGTGAAGTATTTCAAAAAAATATATAAGCTTCATATATT 1920
 DB 1861 GTAAAGACGTGAAGACAGGCTGTGAAGTATTTCAAAAAAATATATAAGCTTCATATATT 1920
 QY 1921 ATTGAATTTCTTGAAGAAATATATACAAATTAATTAACAACACTGGAAGAAATGTCCCTCT 1980
 DB 1921 ATTGAATTTCTTGAAGAAATATATACAAATTAATTAACAACACTGGAAGAAATGTCCCTCT 1980
 QY 1981 TGGCATACGTATTTGCAATATGATTTTCACTTAATTTGTCCAGGCTCTTACGAAGAAATAGT 2040
 DB 1981 TGGCATACGTATTTGCAATATGATTTTCACTTAATTTGTCCAGGCTCTTACGAAGAAATAGT 2040
 QY 2041 TGTCAAAAAATCAAGAAAGCATCGTAGCGCTTCTATTAATAAGAAAGCCCTTGAAGATGCT 2100
 DB 2041 TGTCAAAAAATCAAGAAAGCATCGTAGCGCTTCTATTAATAAGAAAGCCCTTGAAGATGCT 2100
 QY 2101 CTCAAGTAGAGCTTCAAGAAATTTGACTGATTAATCTTAATGTGAACCTGCAATTGAAA 2160
 DB 2101 CTCAAGTAGAGCTTCAAGAAATTTGACTGATTAATCTTAATGTGAACCTGCAATTGAAA 2160
 QY 2161 AGAATTTGTACAGTAGCGGGGAAACGTAAATATGCGTCAATCAATGCGTTATGCAAAAGCT 2220
 DB 2161 AGAATTTGTACAGTAGCGGGGAAACGTAAATATGCGTCAATCAATGCGTTATGCAAAAGCT 2220
 QY 2221 AACACCAAGATTAATCTGGAAGAGTAGAGATGCTAGAAAGAACTGTGAGAAA 2280
 DB 2221 AACACCAAGATTAATCTGGAAGAGTAGAGATGCTAGAAAGAACTGTGAGAAA 2280
 QY 2281 TTAGTGAAGAGTAGAGAAACAGTGCACCAAGCATTTACCAAGAAATTAGACAAACCGCA 2340
 DB 2281 TTAGTGAAGAGTAGAGAAACAGTGCACCAAGCATTTACCAAGAAATTAGACAAACCGCA 2340
 QY 2341 GCTGATCTAAAAAAGATTATTAAGACATATGAGAACTTAAAGAAACGTGCAGAGAAAGCA 2400
 DB 2341 GCTGATCTAAAAAAGATTATTAAGACATATGAGAACTTAAAGAAACGTGCAGAGAAAGCA 2400
 QY 2401 ATGAACAAGTCAGCTCTTTTGTGCACTCACTTAAGAAAAAGAAAGTAATGATCAAAA 2460
 DB 2401 ATGAACAAGTCAGCTCTTTTGTGCACTCACTTAAGAAAAAGAAAGTAATGATCAAAA 2460
 QY 2461 AGTAATAGCAAAAAACAAGATTAAGATGCGCTTCAACGAGCTTCAAGATTAACCAAAA 2520
 DB 2461 AGTAATAGCAAAAAACAAGATTAAGATGCGCTTCAACGAGCTTCAAGATTAACCAAAA 2520
 QY 2521 CATGTGAAAAATCTACGAGGGGAGTTAAGATGTAATCCGTAACGAATTTGAAGCTTAAA 2580
 DB 2521 CATGTGAAAAATCTACGAGGGGAGTTAAGATGTAATCCGTAACGAATTTGAAGCTTAAA 2580
 QY 2581 GCATTTGATTTGGGACGAGAGTATTTTGAAGATTAATGTAATTTGAAGAAAGATGTAAT 2640
 DB 2581 GCATTTGATTTGGGACGAGAGTATTTTGAAGATTAATGTAATTTGAAGAAAGATGTAAT 2640
 QY 2641 AAATTTGGAATCAGATTGCAAGATTAAGAGAGATTGCAAAAGCTTAAAGAAAGTATGCAAA 2700
 DB 2641 AAATTTGGAATCAGATTGCAAGATTAAGAGAGATTGCAAAAGCTTAAAGAAAGTATGCAAA 2700
 QY 2701 AAGATTAAATAGGCTTTGCGCAATCTGAAGCCTCTGAGGTGAAGCCGCAAGAAACAGTG 2760
 DB 2701 AAGATTAAATAGGCTTTGCGCAATCTGAAGCCTCTGAGGTGAAGCCGCAAGAAACAGTG 2760
 QY 2761 ACAGAAAGTACAGAGCAACTTAACAAACAACCAACCGTTGCCGATCCGAAAGCAACG 2820
 DB 2761 ACAGAAAGTACAGAGCAACTTAACAAACAACCAACCGTTGCCGATCCGAAAGCAACG 2820
 QY 2821 GAATGCAAAATCCTTAACAGACAAACAGACATGGTTTACAGACATGCAACACACAAGC 2880
 DB 2821 GAATGCAAAATCCTTAACAGACAAACAGACATGGTTTACAGACATGCAACACACAAGC 2880
 QY 2881 ACGTCTACTATCACTTACCATCACTCAAAAATTAATGATGACATCAACAGGCGAGTGC 2940
 DB 2881 ACGTCTACTATCACTTACCATCACTCAAAAATTAATGATGACATCAACAGGCGAGTGC 2940

QY 2941 AACCAACCAAGTGTACGACAGAGAGAAAGATGATGACAGAGACGTGAACCGAGTGA 3000
 DB 2941 AACCAACCAAGTGTACGACAGAGAGAGAAAGATGATGACAGAGACGTGAACCGAGTGA 3000
 QY 3001 GGGCTGAGATGATGCTGGTGGAAATGTATATGAGGGGGGTATATGACAAATGCTTATTCG 3060
 DB 3001 GGGCTGAGATGATGCTGGTGGAAATGTATATGAGGGGGGTATATGACAAATGCTTATTCG 3060
 QY 3061 TTCATGATTTAG 3072
 DB 3061 TTCATGATTTAG 3072
 RESULT 2
 ID AA294068
 AA294068 standard; DNA; 3054 BP.
 AC AA294068;
 AC 15-SEP-2003 (revised)
 DT 05-JUN-2000 (first entry)
 DE Pneumocystis carinii major surface glycoprotein gene HMSG33.
 KW Major surface glycoprotein; MSG; HMSG33; human; pneumonia; diagnosis; ds.
 OS Pneumocystis carinii; sp. f. hominis.
 XX
 FH Key
 FT 1. .3054
 FT /*tag= a
 FT /product= "HMSG33"
 FT /note= "a nucleic acid comprising residues 2887-3132 of
 this sequence is specifically claimed in Claim 28"
 XX
 PN MO200009760-A1.
 XX
 PD 24-FEB-2000.
 XX
 PF 17-AUG-1999; 99MO-US018750.
 XX
 PF 17-AUG-1998; 98US-0096805P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;
 XX
 DR WPI; 2000-206025/18.
 DR P-PSDB; AAY79170.
 XX
 PT Detection of the presence of Pneumocystis carinii in specimens by
 PT identification of major surface glycoprotein (MSG) gene sequences using
 PT two or more oligonucleotide primers derived from human P. carinii MSG
 PT protein encoding sequence.
 XX
 PS Claim 27; Page 87-91; 110pp; English.
 XX
 CC The present sequence is that of the novel Pneumocystis carinii sp. f.
 CC hominis gene, HMSG33, which encodes a major surface glycoprotein (MSG,
 CC see AAY79170). The gene was isolated by PCR amplification of DNA taken
 CC from an autopsy lung sample of an HIV-infected patient with P. carinii
 CC pneumonia. It is 1 of 7 novel, claimed genes (see AA279063-69) of the
 CC invention that encode human-P. carinii MSGs (see AAY79165-71). The MSGs
 CC include a highly conserved C-terminal region of approximately 100 amino
 CC acids; this region (also claimed) corresponds to residues 2887-3132 of
 CC the present sequence. Direct detection or amplification of human-P.
 CC carinii MSG-encoding genes, especially by PCR using primers directed at
 CC the conserved region of the genes, provides a sensitive and specific
 CC technique for the detection of P. carinii, and the diagnosis of P.
 CC carinii pneumonia, especially in biological specimens (e.g. blood,
 CC sputum) from immunocompromised patients such as those with HIV infection.
 CC (Updated on 15-SEP-2003 to standardise OS field)
 XX

Sequence 3054 BP; 1222 A; 449 C; 670 G; 713 T; 0 U; 0 Other;

Query Match 75.5%; Score 2318.8; DB 3; Length 3054;

Best Local Similarity 86.0%; Pred. No. 0;

Matches 2644; Conservative 0; Mismatches 404; Indels 27; Gaps 6;

```
QY 4 GGGCGGGCGGTCAACCGCAGG---CAGCAGGAGACACGAATGAGATTGAGAGAACAT 60
DB 1 GGGCGGGCGGTCAACCGCAGGAGTAACAGAGCATGAGGCAATATGATGATGAAATGAT 60
QY 61 GTTTAGCTTTTAATCTAAGAGAGATGGACTAAGTGAAGAGATGCAAAAAAACTA 120
DB 61 ATTTGGCGTTGATCTACAGAGAGATGCAATGAGAGATGCAAAATGCAAAAAAGTTTA 120
QY 121 AAAAAATATTTGCCAAGATTTGACTGAGCAAAACTAATATAGAACAACTACAGAAA 180
DB 121 GAAAAATATCTCGAAGAGATGAAAAAGCATCTAGACATGAAAAAGTACATATAATG 180
QY 181 CTTAAGGTTTTTGGGAAGTGGAAAAAGCATACAAAATGCAAAAGAACTGAAAGCCAA 240
DB 181 CTTAAGATTTCTGGAATGGAAGAGCAAGTAAAGCAAAATACAAAATGTCAGGCTCTA 240
QY 241 ATTGAGAAAAAATGTACTACATCAAGAGAAAATTAAAGAACAAATTTAAAAAAAT 300
DB 241 CAAGCCAAAGTTACGGGGAAATGTAACAATTTTAAACCAAAAGCTAGACCAAGCTTA 300
QY 301 CAGATTATACGGAATPAGGATTTGCAAGAGATGAAACAACAATGCTATTTTGGAGGA 360
DB 301 ACAAAATCCATCGAGATGATATTTGCAAGAGAGTGAACGACAATGCTATTTTGGAGGA 360
QY 361 GATATGTTCAAAAGCTTAAGATGATTGCAATCTTTGAGAAATAGTGCATATAAAG 420
DB 361 GATG---CCATTAATCTTTAGAAAGTTGTAACAACTAAGGAATCTAATGTAACAGAAA 417
QY 421 AAACGTGATTAAGTTGCGGAGAGAGTTCTTTTAAGAGCACTTCGAGGATCTTAATGA 480
DB 418 AAACGTGAGGAGTGAAGAGAGAGAGTCTTTTGAAGGACCTTCGATGATCTCAATAAA 477
QY 481 TCAGTCATATGTGAAAAAACTTAAAGAGATTGGCCCTGTCTATGAGGAGGAAGTAT 540
DB 478 ACAGAAACCATGAGAAAAAACTGAAAGAGATTTCGCCAGTCTTCAGAGGGAAGTAT 537
QY 541 GAGTTAACAACTGTGTCTGAACAGAAAGAGACATGAAGATTTTAAATTTGAAGAAA 600
DB 538 GAATTAACGAGCTTGTTGTTGAACAGAAAGAGAGTGCAGAAATTTAATTAAGAAAA 597
QY 601 GATTAAGAGTGGGACTCTTAAAAACAGATGTTTCAAGCAGCACTAGAAATTTAAAAA 660
DB 598 GATTAAGAAATGCACTACTCTTAAAGCAATGTTGCAACAGCACTTGAAGTTTAAAAA 657
QY 661 GAAACATGTTCTTGAATTACTGAAACATGCTAATTTTACATTGGAATTTGCGAGAC 720
DB 658 GAAATATGCTTGAATTACTGAAACATGCTAATTTTACATTGGAATTTGCGAGAC 717
QY 721 GATTAATTAATTAATTAATTAATTTGGAGGAAAAATGCAAGAACAAATATGCTTAATG 780
DB 718 GATTAATTAATTAATTAATTAATTTGGAGGAAAAATGCAAGAACAAATATGCTTAATG 777
QY 781 CCACAGAGACCCGATTTTGAATCAACTAGGCCAGAGGCTACAAATAGCAGAGATTAAGG 840
DB 778 CCACAGAGACCCGATTTTGAATCAACTAGGCCAGAGGCTACAACTAGCAGAGAC 837
QY 841 CTGGAAGAGTTTATTAAGAGGTAGAGAGATGAGATTTTATTTGGAAGAAATCATCTA 900
DB 838 CTGGAATGAGCTTTATTAAGAGGAGAGAGATGAGATTTTATTTGGAAGAAATCATCTA 897
QY 901 AAGAGTGCAGAGCTTTGTTGGCATTTTG---ATCCAGATTTCTAGTCTTAATAAAAAA 957
DB 898 AAGAGTGCAGAGCTTTATTAAGAGGAGATGAGATTTTATTTGGAAGAAATCATCTA 957
QY 958 GAGCAGCAAGAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAAATCTCATGAAC 1017
DB 958 AATATCGAGAAATGCAATGAATCTCGAAGATTAATGCAAAAACTCTCAACAGCAT 1017
```

```
QY 1018 GAGCTTTAGAAAGTTTATGTAAGAAAAATGCTTAAAGTAATGATGAGCAAAAAATGCT 1077
DB 1018 GAGCTCTTAAAAATTTATGTAATTAATTAATGCTTATGATGATGAGCAAAAAATGCT 1077
QY 1078 GAAAGATTTGCAAAATGATATTAACAAACTGCAAAATTTTCACTTCAAAAGTCACTAT 1137
DB 1078 AAGAAATTTAGAGAGATTTATTAAGAAATGATGCAAAACCTCAAAACCAAGATTTTAA 1137
QY 1138 AATGCTCTTTTGTATCCAAACAAAAGAAATATGTAATTTGTTGATGGAAGGCTTGCA 1197
DB 1138 AATCATCTTTATGATCC-----AATGATAAATTTGTTGATGGAAGAAATGCTGCA 1188
QY 1198 ACATTTCTTACGACGAGAGATTGCGCAATTTGAGTCTATTTGTTCTATTTGAAAAA 1257
DB 1189 ACATTTCTTACGATGAGAGATGTCGAGAGATTTGAAATCTTAATTTTCTACGAAAAA 1248
QY 1258 AATGTCAGATGAGAAAAATGATGTAAATTTAAAGAGCAACATGTTACAAAAGAGGA 1317
DB 1249 GCTTGTCAATATGCCAAGAGAGATGTAATTTGAGGCGAGGCTGTTATAGAGAGGG 1308
QY 1318 CTGATGCACGGCAAAATAAAGTCTGCAAGAAAAATGCGAGGAATGTTACATGTTCA 1377
DB 1309 CTTGATGAGCGGCAAAATAAAGTCTGCAAGAAAAATATATGCGTGGTTATATCGTGTCA 1368
QY 1378 AACAAAAGCTGCTTGAAGATTTCAACAGAAATTAATTAAGTAATGTAAGAACTGAAA 1437
DB 1369 AATCAAAAGTTGCTTAAGAGATTTCAACAAAGATTAATTAAGTAATGTAAGAACT--A 1425
QY 1438 AAGAAAAAAGAGAGTTTCTCAAGCATGAAATTAATTTATTTCTGTGTGTCAGCCAGCA 1497
DB 1426 AAGAAAAATTAAGAGAGTTTCCCAAGCATGAATATTTGTTCTGTGTGTCAGCCAGCA 1485
QY 1498 AAAGCAGCCGTTGCTTACATGATCTTGAATGAAGAACTATCTTTTACGACAA 1557
DB 1486 AAAGCTGACGATTAATTAACAGATCATCAATGAGGTTATCTTTTACGACAA 1545
QY 1558 CTGATCAAAAGCGAGATTTCCCGACAGATTAATAATTTGCAAGAAATTTGGGAGAAAGTGC 1617
DB 1546 TTGATCAAAAGAGAGATTTCCGACAGATTAACATGCAAGAGAAATTTGGGAGAAATGTC 1605
QY 1618 CAAGATTAAGAGAGATTTCAAAAGAAATTTACATGGCCATGCTATCACTGAGAGCA 1677
DB 1606 CAAGATTTAGAGAGATTTCAAAAGAAATTTACATGGCCATGCTATCACTGAGAGCA 1665
QY 1678 TGCAATGCTTTGGGACTACAGAAATTTTAAAGCAGTTTATTTGATGAACACAAAGAT 1737
DB 1666 TGCAATGCTTTGGGACTACAGAAATTTTAAAGCAGTTTATTTGATGAACACAAAGAT 1725
QY 1738 ACTTTGAAGACCAAGAAATTTGTAATAATCTTAAAGAAAGAGTGAATTAATGCTCT 1797
DB 1726 ACTTTGAAGACCAAGAAATTTGTAATAATCTTAAAGAAAGAGTGAATTAATGCTCT 1785
QY 1798 AGAAGAGAGATGACCGTTCTCTTTTGTATGTCTTCAAAACGCTACAGTGTGAGCTG 1857
DB 1786 AGAAGAGAGATGACCGTTCTCTTTTGTATGTCTTCAAAACGCTACAGTGTGAGCTG 1845
QY 1858 ATGCTAAAGACGTGAAGACAGGTGTGAATTTCAAAAAATATTAAGCTTCATAT 1917
DB 1846 ATGCTAAAGACGTGAAGACAGGTGTGAATTTCAAAAAATATTAAGCTTCATAT 1905
QY 1918 ATTAATGAATTTCTTGAATAATTAATCAATTAATAACACATGGAAGAAATTTGCTCC 1977
DB 1906 ATTAATGAATTTCTTGAATAATTAATCAATTAATAACACATGGAAGAAATTTGCTCC 1965
QY 1978 TCTTGGCATAGATTTGCAATAGATTTTCACTAATTTGTCAGAGTCTTACAGAAAGAGAT 2037
DB 1966 TCTTGGCATAGATTTGCAATAGATTTTCACTAATTTGTCAGAGTCTTACAGAAAGAGAT 2025
QY 2038 AGTTGTCAAAAAATCAAGAACATGTGAGCCGTTCTATAAAGAGAGCCCTTGAGAGAT 2097
DB 2026 AGTTGTCAAAAAATCAAGAACATTTGTGAGCCGTTCTATAAAGAGAGCCCTTGAGAGAT 2085
```

Qy	2098	GCCTCAAGTAGAGCTTCAAGGAAATTTGACTATTAATCTAAATGGAACCTGCATTCG	2157
Db	2086	GCTTCCAAAGTAGAGCTTCAAGGAAATTTGACTATTAATCTAAATGGAACCTGCATTCG	2145
Qy	2158	AAAAGATATTGTACAGTAGCCGGAAACCTAAATATAGCGTCAATCAGTGGCTTATGCAAA	2217
Db	2146	AAAAGATATTGTACAGTAGCCGGAAACCTAAATATAGCGTCAATCAGTGGCTTATGCAAA	2205
Qy	2218	GCTAACCAACAGGATTAACCTCTGGAAAAGGTGATGAGATGCTTAGAAAAGAACTCTGTGAG	2277
Db	2206	GCTAACCAACAGGATTAACCTCTGGAAAAGGTGATGAGATGCTTAGAAAAGAACTCTGTGAG	2265
Qy	2278	AAATTAGTGAAGAAGTGGAAAGAAACAGTGCAAAGATTACCAACAGAAATTAGACAAACG	2337
Db	2266	AAATTAGTGAAGAAGTGGAAAGAAACAGTGCAAAGATTACCAACAGAAATTAGACAAACG	2325
Qy	2338	GCAAGCTGATCTTAAAAAAAAGATTATPAGA CATATGAGAACTTTAAGAAAAGTGCAGAGAA	2397
Db	2326	GCAAGCTGATTTAAAAAAAAGATTATPAGA CATATGAGAACTTTAAGAAAAGTGCAGAGAA	2385
Qy	2398	GCAATGAACAAGTCCAGCTCTTGTGTTTGTGCATCTTAAGAAAAACGAAAGTAAATGTATCA	2457
Db	2386	GCAATGAACAAGTCCAGCTCTTGTGTTTGTGCATCTTAAGAAAAACGAAAGTAAATGTATCA	2445
Qy	2458	AAAAGTAATAGCAAAAAACAAGATPAGAATCCGTTTCAACCGGACTTCAAGATACCA	2517
Db	2446	AAAAGTAATAGCAAAAAACAAGATPAGAATCCGTTTCAACCGGACTTCAAGATPACCA	2505
Qy	2518	AAACATGTGAATAATACTACGAGGGGAGTTAAGATGTATCCGTAAACAAATTAGAACCT	2577
Db	2506	AAACATGTGAATAATACTACGAGAGAGAGTTAAGATGTATCCGTAAACAAATTAGAACCT	2565
Qy	2578	AAAGCATTTGATTTGGCAGCAAGAAATATTGGAAATATGTAGATTTGAAGAAAGATGT	2637
Db	2566	AAAGCATTTGATTTGGCAGCAAGAAATATTGGAAATATGTAGATTTGAAGAAAGATGT	2625
Qy	2638	AATTAATTTGGAAATCAGATTGCAGAATTTAAGAGAGATTGCAAAAGACTTGAAGAAAGTATGC	2697
Db	2626	AATTAATTTGGAAATCAGATTGCAGAATTTAAGAGAGATTGCAAAAGACTTGAAGAAAGTATGC	2685
Qy	2698	AAAAAGATTAATAAGGCTTGTGCAATCTGAAGCCTCTGAGGTGAAACCCGACGAAACA	2757
Db	2686	AAAAAGATTAATAAGGCTTGTGCAATCTGAAGCCTCTGAGGTGAAACCCGACGAAACA	2745
Qy	2758	GTGACAGAAACTACAAACGAACTPACAAACAACAACCGTGGCCGATCCGAAAGGCA	2817
Db	2746	GTGACAGAAACTACAAACGAACTPACAAACAACAACCGTGGCCGATCCGAAAGGCA	2805
Qy	2818	ACGGAATGCAAAATCCTTACAGACAAACAGACACATGGGTTTACACAGACATCGACACACA	2877
Db	2806	ACGGAATGCAAAATCCTTACAGACAAACAGACACACATGGGTTTACACAGACATCGACACACA	2866
Qy	2878	AGCAAGCTTACTATCAATCTTACCATCAATCAAAAAATPACATTTGACATCAACGAGCGCA	2937
Db	2866	AGCAAGCTTACTATCAATCTTACCATCAATCAAAAAATPACATTTGACATCAACGAGCGCA	2925
Qy	2938	TGCAAAACCAACCAAGTGTACGACGGAAGGAAGTATGCAAGGAGAGTGAACCCGAGT	2997
Db	2926	TGCAAAACCAACCAAGTGTACGACGAG-----GAGTATGCAAGGAAGAGTGAACCCAGT	2979
Qy	2998	GAGGGGCTGAGATGAGTGGGTGGAATGTGATGAGGGGGGTGATGTAGCAATGGTTATT	3057
Db	2980	GAAAGGCTTGAAGGGTGAAGCGGGGTGGAATGTGATGAGGGGGGTGATGTAGCAATGGTTATT	3033
Qy	3058	TCGTTTCATGATTTAG 3072	
Db	3040	TCGTTTCATGATTTAG 3054	

RESULT 3
AAZ94066
ID AAZ94066 Standard; DNA; 3084 BP
XX

AC	AA294066;
XX	
DT	15-SEP-2003 (revised)
DT	05-JUN-2000 (first entry)
XX	
XX	
DE	Pneumocystis carinii major surface glycoprotein gene HMSG14.
XX	
KM	Major surface glycoprotein; MSG; HMSG14; human; pneumonia; diagnosis; ds.
OS	
XX	Pneumocystis carinii; sp. f. hominis.
XX	
FH	
FT	Key Location/Qualifiers
FT	1..3084
FT	/*tag=
FT	/product= "HMSG14"
FT	/note= "a nucleic acid comprising residues 2839-3084 of
FT	this sequence is specifically claimed in Claim 28"
PN	WO200009760-A1.
XX	
PD	24-FEB-2000.
XX	
PF	17-AUG-1999; 99WO-US018750.
XX	
PR	17-AUG-1998; 98US-0096805P.
XX	
XX	(USSH) US DEPT HEALTH & HUMAN SERVICES.
PA	
PI	Kovacs JA, Huang S, Masur H, Fischer SH, Gyll VJ, Mei Q;
DR	WPI; 2000-206025/18.
XX	P-PSDB; AA79168.
XX	
PT	Detection of the presence of Pneumocystis carinii in specimens by
PT	identification of major surface glycoprotein (MSG) gene sequences using
PT	two or more oligonucleotide primers derived from human P. carinii MSG
PT	protein encoding sequence.
XX	
PS	Claim 27; Page 72-76; 110pp; English.
XX	
CC	The present sequence is that of the novel Pneumocystis carinii sp. f.
CC	hominis gene, HMSG14, which encodes a major surface glycoprotein (MSG,
CC	see AA79168). The gene was isolated by PCR amplification of DNA taken
CC	from an autopsy lung sample of an HIV-infected patient with P. carinii
CC	pneumonia. It is 1 of 7 novel, claimed genes (see AA79063-69) of the
CC	invention that encode human-P. carinii MSGs (see AA79165-71). The MSGs
CC	include a highly conserved C-terminal region of approximately 100 amino
CC	acids; this region (also claimed) corresponds to residues 2839-3084 of
CC	the present sequence. Direct detection or amplification of human-P.
CC	carinii MSG-encoding genes, especially by PCR using primers directed at
CC	the conserved region of the genes, provides a sensitive and specific
CC	technique for the detection of P. carinii, and the diagnosis of P.
CC	carinii pneumonia, especially in biological specimens (e.g. blood,
CC	sputum) from immunocompromised patients such as those with HIV infection.
CC	(Updated on 15-SEP-2003 to standardise OS field)
XX	
SQ	Sequence 3084 BP; 1240 A; 455 C; 676 G; 713 T; 0 U; 0 Other;
XX	
Query Match	73.7%; Score 2262.6; DB 3; Length 3084;
Best Local Similarity	84.7%; Pred. No. 0; Mismatches 399; Indels 78; Gaps 6
Matches 2840; Conservative	0; Mismatches 399; Indels 78; Gaps 6
DB	1 ATGGCGCGGCGGTCAAGCGGCAGGACAGGACACAGAAATGACATTGATGAGAACAT 60
QY	1 ATGGCGCGGCGGTCAAGCGGCAGGACAGGACACAGAAATGACATTGATGAGAACAT 60
DB	1 ATGGCGCGGCGGTCAAGCGGCAGGACAGGACACAGAAATGACATTGATGAGAACAT 60
QY	61 GTTTAGCTTTAATCTTAAGAAAGATGACTAAGTGACGAGAAATGCAAAAACTA 120
DB	61 GTTTAGCTTTAATCTTAAGAAAGATGACTAAGTGACGAGAAATGCAAAAACTA 120
QY	121 AAAAATTTTCCCAAGATTGACTGCAAGCAAACTAAATATAGAACAAAGTACACAGAAA 180
DB	121 AAAAATTTTCCCAAGATTGACTGCAAGCAAACTAAATATAGAACAAAGTACACAGAAA 180
QY	121 GAAAGATTTTCCCAAAACCTTAAACAATGACAGATTAAATTCAGAAAAGTTTCAAGAAA 180
DB	121 GAAAGATTTTCCCAAAACCTTAAACAATGACAGATTAAATTCAGAAAAGTTTCAAGAAA 180

OY 181 CTTAAGGTTTTTGGGAAGTGGAAAGCATACAAAATGCAAAAGAACTGAAACCAAT 240
DB 181 TTTAAAGATTTCTGTGATACCGGAAACGAAATGAAAAATGCAAGATCAAAAAACAA 240
OY 241 AATTGAAAAAATGACTACATCAAGGAAAACTTAAAGAAAGCAATTAAAAAAATTT 300
DB 241 GTCATCAAAAATGCAATTAATTTCAAGGAAAACTTCAACAGCTGCTAGAAAAAAATTT 300
OY 301 CAGATTATACGAGATAGGATTTGCAAGAGAAATGAAACAACATGCTATTTTGGAGGA 360
DB 301 TCAGAAATTAACAGATGAGGATTTGCAAAAAGAAATGAAACAATGCTATTTTGGAGGA 360
OY 361 GTATGTTCAAAAGAACTTAAGATGATTGCAATCTTTGAGAAATAGTCTATCAAAAG 420
DB 361 GCATGTCACAACGAACCTTAAGATGACTGCAATTAATTAGGAAATACGTTATCAAAA 420
OY 421 AAAGGTGTAAGGTGGGGAAGAGTTCTTTAAAGAGCACTTCGAGGATCTTAAATGA 480
DB 421 GAAAGGAAACAATGTGGCAAGAAAGTTCTTTGAGGGGCGCTTCGTGATCTCAATGA 480
OY 481 TCAGTCATATGTGAAAAAAAACCTTAAGAGATTTGCCCTGTCATGGGAGGAAAGTAT 540
DB 481 ACAAGACATGTGAAAAAAAACCTTAAGAGATTTGCCCGAAATTGAAAGAGAAAGCAT 540
OY 541 GAGTTAACAAACTTGTGTCTGAACCAAGACAGACATGTAGAAATATTTTAAATGA 600
DB 541 GAATTTACGAGCTTGTCTTATCAAAAACACATGCGTAAAGCTTGTAAACAAAGGA 600
OY 601 GATTAAGAGTGGCGTACTCTTAAACAGATGTTTACGAGCACTAGGAAGTTTAA 657
DB 601 AAAATTAATGTGATCTCTTAAAGAAAGTTGAAGAGCACTTAAGAGATTAATTTG 660
OY 658 AAAGAAACATGTCTTGAATTACTCGAACATGCTATTTTACATTGGAATTTGCCAG 715
DB 658 CGAGAAAAATGTCTACTATTACTTAGAGCAATGTTACTTCAAGAGGAACTGTAAAGA 720
OY 716 -----ACGACGATTAATAATGT 735
DB 721 GACAAATCAAAAGTGCATTAACCTTAATAATAAGACTGCAAGAAATATGTACAGAGTGT 780
OY 726 AATTGAATGGGAGAAAAATGCAAGAACAAATATTTGCTTATATGCAACGAGACCGAT 795
DB 781 GATGAATTAAGCAAAAAGTGTGAAAAAGAAATATTTGTTATATCATCCAGATCCGAT 840
OY 796 TTTGATCCAACTAGGCGCAAGGCTCAATAGCAGAGATATAGGGCTGAAAGATTTTAT 855
DB 841 TTCGATCCAACTAGGCGCAAGGCTCACTAGCAGAGAGCATAGGGCTGAAAGACCTTAT 900
OY 856 AAAGAGGTAGAGAGAGATGGATTTTATTTGAAAGAAATCATCTAAGAGTGCACAGCT 915
DB 901 AAGAGGGCAGAGAGAGATGGATTTTGTGTGAAACAACATGTAAAGATGCAACACT 960
OY 916 TTGTTGGCATTGTTATCCAAAGATCTAGTCTTAAAAAAAAGACGACGAAGAGAAATGC 975
DB 961 TTGTTGGCATTACT-----TCTTAAAGAAACCTTTAAAAAAGAGAAAGTGT 1005
OY 976 GAAGAGCCCTTCAAAAAAGCTGCAAAAATCCTCATGAAACATGAGGCTTTAGAAAGTTTA 1035
DB 1006 ATAAAGCCCTTAAAAAAAAGCTGCAAAAACCTCATGAAACATGAGGCTTTAGAAATGTA 1065
OY 1036 TGTAAAGAAATGGTTTAAGTAAATGATGAAACGAAAAAATGTGAAGAAATTTGCAAAATGAT 1095
DB 1066 TGTAAAGAAATTAACAAAGTATGATGAAACGAAAAAATGTGAAGAACTAGAAAAAATAT 1125
OY 1096 ATTAACAAAACCTTGCAAAATTTTCACTTCAAAAGTCACATATATGCTTTTGTATCA 1155
DB 1126 GTTAAACAAAACCTTGTACAGTCTTACATCAACATTTCTTAAAAACGCTCTTTACATTTCA 1185
OY 1156 ACAAAGAAATATGAATTTGTTGATGGGAAGGTTGCCAACTTTCTTAGCAACGAA 1215
DB 1186 CC-----TGATGAATTCGGAAATGGGGAATAATTAACGACATTTCTTAGTATGA 1236

OY 1216 GATTGCGAAATTTGAGTCTATTTGTTCTATTTTGAAAAAAAATGTCAGATGAGAA 1275
DB 1237 GATTGCGAAACCTAGATCTTATTTGCTTTTATTAAGAACTTGTCCGATGTCAA 1296
OY 1276 AATGATGTAAAAATTAAGAGCAACATGTTACAAAAGAGACTTGATGACGGGCANAAT 1335
DB 1297 GAACTGTATGAATGTGAGGGCAGCGTGTTCAGAGAGGGCTTGATGACGGGCAAAAC 1356
OY 1336 AAAGTGTGCAAGAAAAATGCGAGGAATGTTACATGGTTCAAAACAAAGCTGGCTTGA 1395
DB 1357 AGTGTGTTGCAAAAATATATGCGTGGTTATTAAGTGGTCAAAATGAAAGTTGCTTAAG 1416
OY 1396 AAGTTTCAACAAGAAATTAAGTAAAGTATGAGAAACGAAAAAAGAAACAAAGAGAT 1455
DB 1417 GAGTTTCAACAAGAAATTAAGTAAAGTATGTAAGAGCT--AAAGAAAAATTAAGAGAGT 1473
OY 1456 TTCTCAACGATGAATTAATTAATCTGTGATGACGCAAGCAAAAGCGCGGTGCTT 1515
DB 1474 TTCCCAACGATGAATTAATTTGTTCTGTGTACAGCCAGCAAAAAGCTGCACGATTACTT 1533
OY 1516 ACACATGATCTTGCATGAAAACTATCTTTTACGACAAACAATGATCAAAAGCGAGAT 1575
DB 1534 ACACACGATCATCAATGAGGTTACCTTTTACGACAAACAATGATCAAAAGAGAGAT 1593
OY 1576 TTCCCGACAGTAAAAATTTGCAAGGAATGGGGGAAAAAGTCCCAAGATTTAGAGAGAT 1635
DB 1594 TTTCGACAGATTAAGACTGCAAGGAACTAGGAAAAAATGCCAAAGATTTAGAAAGAGT 1653
OY 1636 TCAAAAGAAATTAACATGGCCATGTCATACATGAGCAGCAATGCAATCGCTTGGGACT 1695
DB 1654 TCAAAAGAAATTAACATGGCCATGTCATACATGAGCAGCAATGCAATCGCTTGGGACT 1713
OY 1696 ACAGAAATTTTAAAGCAGGTTTATTTGATGAACACAAAGATCTTTGAAAGCAAGAA 1755
DB 1714 ACAGAAATTTTAAAGCAGGTTTATTTGATGAACACAAAGATCTTTGAAAGCAAGAA 1773
OY 1756 AGTGTGTAATAATCTTAAAGAAAGATGTAAATAGTCTTAAAGAGAGATGACCGT 1815
DB 1774 AGTGTGTAATAATCTTAAAGAAAGATGTAAATAGTCTTAAAGAGAGATGACCGT 1833
OY 1816 TTCTCTTTGATGTGCTTCAAAAACGCTACGCTGAGCTGATGTAAAGAGCTGAAA 1875
DB 1834 TTCTCTTTGATGTGCTTCAAAAACGCTACGCTGAGCTGATGTAAAGAGAGCTGAAA 1893
OY 1876 GACAGGTGTGAATTTCAAAAAAATTAAGACCTTCATATATTAATTTCTTGAA 1935
DB 1894 GACAGGTGTGAATTTCAAAAAAATTAAGACCTTCATATTAATTTCTTGAA 1953
OY 1936 AATAATCAAAATAAATAACAAACCTGAAAGAAATTTGCTTGGCAATACGATATTC 1995
DB 1954 AATAATCAAAATAAATAACAAACCTGAAAGAAATTTGCTTGGCAATACGATATTC 2013
OY 1996 AATGATTTTCACTTAATTTGTCAGAGTCTTACGAAAGAAATTAAGTGTACAAAAATCAAG 2055
DB 2014 AATGATTTTCACTTAATTTGTCAGAGCTTACGAAAGAAATTAAGTGTGTACAAAAATCAAG 2073
OY 2056 AAGCATGCTGAGCGCTTCTATTAAGAAAGAGCGCTTGAAGATGCTTCAAGTGAAGCTT 2115
DB 2074 AAGCATGCTGAGCGCTTCTATTAAGAAAGAGCGCTTGAAGATGCTTCAAGTGAAGCTT 2133
OY 2116 CAAGAAAAATTTGATGATTAATCTTAATGTGAACCTGATGAAAGAAATTAATTTGACGTA 2175
DB 2134 CAAGAAAAATTTGATGATTAATCTTAATGTGAACCTGATGAAACAGATATTTGACGTA 2193
OY 2176 GCGGGAACGTAATTAATGCGTCAATCAATGAGGCTTATAGCAAAAGCTTACCAAGATTAAC 2235
DB 2194 GCGGGAACGTAATTAATGCGTCAATCAATGAGGCTTATAGCAAAAGCTTACCAAGATTAAC 2253
OY 2236 TCTGAAAGAGTATGAGATGCTAGAAAGGAACCTGTGTGAATTAAGTGAAGAAAGT 2295
DB 2254 TCTGAAAGAGTATGAGATGCTAGAAAGGAACCTGTGTGAATTAAGTGAAGAAAGT 2313
OY 2296 GAAGAACGTGCAAAAGCATTTACCAACGAATTAAGACAAACGGGACGCTGATCTAAAAAA 2355

```

Db      2314  GAAAGACAGTCGAAGCATTTACCAACAGAAATTAGACAAACCGGACGCTGATCTAAAAAA 2373
Qy      2356  GATTATAGACATATGAGGAACCTTAAGAAACGTGACAGAGAAAGCAACCAATCCAGT 2415
Db      2374  GATTATAGACATATGAGGAACCTTAAGAAACGTGACAGAGAAAGCAACCAATCCAGT 2433
Qy      2416  CTTGTTTGTCTCATTTAAGAAAAAGAAAGTATGATCAAAAAGTAAATAGCAAAAAC 2475
Db      2434  CTTGTTTGTCTCATTTAAGAAAAAGAAAGTATGATCAAAAAGTAAATAGCAAAAAC 2493
Qy      2476  AAGATTAAGAAATGCGCTTTCAAACGGAAGTCAAGATACCAAAAAACATGTGAAAACTA 2535
Db      2494  AAGATTAAGAAATGCGCTTTCAAACGGAAGTCAAGATACCAAAAAACATGTGAAAACTA 2553
Qy      2536  CGGAGGGGAGTAAAGATGATTCGCTAACAATTAAGAAAGTAAAGATTTGATTTGGCA 2595
Db      2554  CGGAGGGGAGTAAAGATGATTCGCTAACAATTAAGAAAGTAAAGATTTGATTTGGCA 2613
Qy      2596  GCAGAAATATTTGGAAGATATGATTTGAAGAAAGATTAATTAATTTGAATCAGAT 2655
Db      2614  GCAGAAATATTTGGAAGATATGATTTGAAGAAAGATTAATTAATTTGAATCAGAT 2673
Qy      2656  TGCAGAAATTAAGAGATTGCAAAAGCTTAGAAGAAAGTATGCAAAAAGATTAATAGGCT 2715
Db      2674  TGCAGAAATTAAGAGATTGCAAAAGCTTAGAAGAAAGTATGCAAAAAGATTAATAGGCT 2733
Qy      2716  TGTGCGCAATCTGAAGCCCTCTGAGAGTGAAGCCGCAACAAAGTGAACAAAGTACAAAG 2775
Db      2734  TGTGCGCAATCTGAAGCCCTCTGAGAGTGAAGCCGCAACAAAGTGAACAAAGTACAAAG 2793
Qy      2776  ACAACTACAACAACAACAACACCGTTGCCGATCCGAAGGCAAGCAATGCAATCCCTTA 2835
Db      2794  ACAACTACAACAACAACAACACCGTTGCCGATCCGAAGGCAAGCAATGCAATCCCTTA 2853
Qy      2836  CAGACAAACAGACACATGCTGTTACACAGACATGCAACACACACACACACACACACACAC 2895
Db      2854  CAGACAAACAGACACATGCTGTTACACAGACATGCAACACACACACACACACACACACAC 2913
Qy      2896  TCTTACCATCATCAATCAATTAATCAATTCATCAACAGAGCGATGCAACCAACCAAGTGT 2955
Db      2914  TCTTACCATCATCAATCAATTAATCAATTCATCAACAGAGCGATGCAACCAACCAAGTGT 2973
Qy      2956  ACGACAGAGAGAAATGATGATGACAGAGACGTGAACCGAAGTGAAGGCTGAGATGAGT 3015
Db      2974  ACGACAGAG-----GATGATGACAGAGACGTGAACCGAAGTGAAGGCTGAGATGAGT 3027
Qy      3016  GGGTGAATGATGAGAGGGGGGTATGATGCAATGCTTATTTGCTTCATGATTTAG 3072
Db      3028  GGGTGAATGATGAGAGGGGGGTATGATGCAATGCTTATTTGCTTCATGATTTAG 3084

RESULT 4
AA294065 ID AA294065 standard; DNA; 3090 BP.
XX
AC AA294065;
XX
XX 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX
DE Pneumocystis carinii major surface glycoprotein gene HMSG11.
XX
KM Major surface glycoprotein; MSG; HMSG11; human; pneumonia; diagnosis; ds.
XX
XX Pneumocystis carinii; sp. f. hominis.
OS
XX Key Location/Qualifiers
FH 1. 3090
FT CDS
FT
FT /*tag= a
FT /product= "HMSG11"
FT /note= "a nucleic acid comprising residues 2845-3090 of
FT this sequence is specifically claimed in Claim 28"

```

```

XX      XX      WO200009760-A1.
XX      FN
XX      PD 24-FEB-2000.
XX      PD
XX      PF 17-AUG-1999; 99WO-US018750.
XX      PR 17-AUG-1998; 98US-0096805P.
XX      PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX      PI Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;
XX      DR WPI; 2000-206025/18.
XX      DR P-PSDB; AAY79167.
XX      PT Detection of the presence of Pneumocystis carinii in specimens by
XX      PT identification of major surface glycoprotein (MSG) gene sequences using
XX      PT two or more oligonucleotide primers derived from human P. carinii MSG
XX      PT protein encoding sequence.
XX
XX      PS Claim 27; Page 64-68; 110pp; English.
XX
XX      CC The present sequence is that of the novel Pneumocystis carinii sp. f.
XX      CC hominis gene, HMSG11, which encodes a major surface glycoprotein (MSG,
XX      CC see AAY79167). The gene was isolated by PCR amplification of DNA taken
XX      CC from an autopsy lung sample of an HIV-infected patient with P. carinii
XX      CC pneumonia. It is 1 of 7 novel, claimed genes (see AAY7963-69) of the
XX      CC invention that encode human-P. carinii MSGs (see AAY79165-71). The MSGs
XX      CC include a highly conserved C-terminal region of approximately 100 amino
XX      CC acids; this region (also claimed) corresponds to residues 2845-3090 of
XX      CC the present sequence. Direct detection or amplification of human-P.
XX      CC carinii MSG-encoding genes, especially by PCR using primers directed at
XX      CC the conserved region of the genes, provides a sensitive and specific
XX      CC technique for the detection of P. carinii, and the diagnosis of P.
XX      CC carinii pneumonia, especially in biological specimens (e.g. blood,
XX      CC sputum) from immunocompromised patients such as those with HIV infection.
XX      CC (Updated on 15-SEP-2003 to standardise OS field)
XX
XX      SQ Sequence 3090 BP; 1255 A; 447 C; 662 G; 726 T; 0 U; 0 Other;
XX
XX      Query Match 61.9%; Score 1900.6; DB 3; Length 3090;
XX      Best Local Similarity 77.8%; Pred. No. 0;
XX      Matches 2430; Conservative 0; Mismatches 609; Indels 84; Gaps 8;
XX
Qy      1 ATGGCGGCGGCGGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
Db      1 ATGGCGGCGGCGGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
Qy      61 GTTTAGCTTTAATTTCAAGAAAGATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 120
Db      61 GTTTAGCTTTAATTTCAAGAAAGATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 120
Qy      121 AAAAAATATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAG 180
Db      121 GAAGAAATATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAG 180
Qy      181 CTTAAGGTTTTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAG 240
Db      181 TTAAGAAATTTCTGTGTAACGCGGAAACGAAATGAAATGTCAGATCTTAAGAAACAAA 240
Qy      241 ATTGAGAAAAATGTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 300
Db      241 GTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 300
Qy      301 CAGATTAAACGATTAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAG 360
Db      301 TCAGAAATTAACAGATTAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAG 360
Qy      361 GTATGTTCAAAAGAACTTAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAG 420
Db      361 GCATGTTCAACAGAACTTAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAGATTCGCAAG 420

```

OY	421	AAAGTGATTAAGTTGGCGAAGAA	GTTCTTTTAAAGACACTTGGTACGATCTTAATGCA	480
Db	421	GAAGGAAACAATGAGCGAAGAA	AGTTCTTTTGGAGGGCTTCGTGTGATCTCAAGAA	480
OY	481	TCAGTCATATGTGAAAAAAA	CTTAAGAGATTTGGCCCTGTCAATGGGAGGGAATGAT	540
Db	481	ACAAAGACATGTGAAAAAAA	AGCTGAAGAGTTTGCCGAATTGAAGAAGAAAGCGAT	540
OY	541	GAGTTAACAACTTGCTGTGA	ACCAAGACATGTGAAGAAATATTTTAATTTGAAAAA	600
Db	541	GAATTAACGAGGCTTTGTCTTTAT	CAAAAAACAATGCGTAAGCTTTGTAACAAAAGGA	600
OY	601	GATTAAGAAATGCGGTA	CTTTAAAAACAGATGTTTCAAGACACTAAGAACTTTTAA---	657
Db	601	AAAAGTAATGTGATACCTCTTG	AAAAAGAAAGTGAAGACCTTAAGAAAGAAATGAATTG	660
OY	658	AAAGAAACATGTCTTGAATTA	CTGGAACAATGCTAATTTTTCATTTGGAAATTTGGGAG--	715
Db	661	CGAGAAAAATGTCTA	CTATTACTTGAGCAATGTTACTTTCACAAGGGAACTGTGAAGGA	720
OY	716	-----	-----ACGACGATATATTAATTAATGT	735
Db	721	GACAAATCAAAAGTGC	AAATAAATAAAGCTGCACAAAGAAATGTATGACAGATGT	780
OY	736	ATTGTAATTTGGAGGAAAA	ATTCGAACAATAATTTGCTTAATATGCCACAGACCCGAT	795
Db	781	GATGAATTTAGCAGAAA	AGTGTGGAATAAATAATTTGTTTATATGCAATCCAGATCCGAT	840
OY	796	TTTAATCCAACTAGGCG	CAGAGCTACATATACAGAGATATAGGGCTGGAAAGATTTAT	855
Db	841	TTGATCCAACTTAAGCC	AGACTACACTACAGAGCATATAGGAGCTGGAAAGACTTTAT	900
OY	856	AAGAAGTATAGAGAGAT	TGAGATTTTATTTGGAAGAATCATCTAAGAGATGCGACACT	915
Db	901	AAAGGGCAGAAAGAGAT	TGGAATTTTGTGGAAGACAATGTATAGAGATGCAACAGCT	960
OY	916	TTGTTGGCATTTGTGAT	CCAAAGATTTCTTAATAAAAAAAGCAGCAAAAGAGAAATGC	975
Db	961	TTGTTGGCACA	TACT-----TCTTAAGAAAACCTTTAAAAAAGAAAGAAATGT	1005
OY	976	GAAGAAGCCCTTCA	AAAAAGCTGCAAAATCCTCATGAACATGAGGCTTTTGAAGTTTA	1035
Db	1006	ATAAAGCCCTTA	AAAAAAACCTGCAAAACCTCATGAACATGAGGCTTTGAAAAATCTA	1065
OY	1036	TGTAAGAAAAATG	TTTAAGTATATGATGGAACGAAAAAATGTGAAGATTCAAAAATGAT	1095
Db	1066	TGTAAAGAAAAAT	TAAACCAAGTATGTATGGAACGAAAAAATGTGATGAATGAAAAAGAT	1125
OY	1096	ATTAACAACA	CTTGCAAAATTTTCACTTCAAAAGTCACTAATTAATCGTCTTTTGTATCCA	1155
Db	1126	GTTAACAA	AACTGTATACAAAGTCTTAACATACTTAATAAAACCGCTTTACATTTCA	1185
OY	1156	ACAAAAGAAAT	TATATGAATTTGTTGATGGGAAGGTTGCCAAGCTTTCTTACGAAGAA	1215
Db	1186	CC-----	TGATGGAATTTGCGAAATGGGAAAAATTAACCGACATTTCTTAATGATGAA	1236
OY	1216	GATTTGCGAATTTG	AGATCTCTATTTGTTCTAATTTTGAAAAAAATGTCAGATGAGGAA	1275
Db	1237	GATTTGCAAA	CTTATGATCTTTATTTGCTTTTATTTATTAAGAAACTGTGCCAGATGTCAA	1296
OY	1276	AATGATGTAAAA	ATATTAAGACAATGTTTACAAAAGAGACCTTGATGACCGGGCAAT	1335
Db	1297	GAAAGCTTAT	ATGATGAGGCGAGCGTGTATTAAGAGAGGCTTGATGACCGGGCAAAAC	1356
OY	1336	AAAGTCTGCA	AGAAAAATATGCGAGGAATGTTACATGTTTCAACAAAGCTGGCTTGA	1395
Db	1357	AGTGTGTGCA	AAAAAATATGCGAGGTTATTTGCAATGCTCAAAATTAAGATTTGACTTAAG	1416
OY	1396	AAAGTTTCA	ACAAGATTTAGTAAAGTATGTGGAACGTGAAGAAAAAGAAAACAAAAGAGAT	1455
Db	1417	AAATTTCA	ACAAGATTTTGAACAAGTATGTGAGAACTG---AAGGAAAAATTAAGAGAT	1473
OY	1456	TTCTTCA	ACGATTAATTAATTTCTGTGTATACAGCCAGCAAAAGCACCCGGTGTCTT	1515

Db	1474	TTCTCGAACGATGATTTGTTGTCTGTGTAACACACGAAAGCAGACGATTTACTT	1533
OY	1516	ACACATGATCTTCGAATGAAAACTATCTTTTTCGACAAACACTGATCAAAAGCGAGAT	1575
Db	1534	ACACATCAACCATCAAAATGAGAGTTATCTTTTTCGACAAACACTGATCAAAAGAGAT	1593
OY	1576	TTCCCGACAGATTAATAAATTCGAAAGAAATGGGGGAAAGTCCCAAGATTTTACGAGAGAT	1635
Db	1594	TTTCCGACAGATTAAGACTCGACAGAAATTTGGGGGAAAGTCCCAAGATTTTACGAAAGAT	1653
OY	1636	TCAAAGAAATTTACATGGCAATGCAACACTGAGACGAATGCAATCCCTTGGGGACT	1695
Db	1654	TCAAAGAAATTTACATGGCCATGCAACACTGACACCTGAACAGCAATGCAATCCCTTGGGATTT	1713
OY	1696	ACAGAAATTTTAAAGCAGTTTTTTTGGATGAAACAAAGATACTTTGAAAGACCAAGAA	1755
Db	1714	ACAGAAATTTTAAACAGATTTTATTTGATGAAACAAAGATACTTTGAAAGATCAAGAA	1773
OY	1756	AGTTGTGTAATAATCCTPAAAGAAAGTGTATTAATGTGCTAGAGAGAGATGACCGT	1815
Db	1774	AACGTGCAAAAATTTTAAAGAAATGCCATTAATGTGCTAGAGAGGGGTGATGATCGT	1833
OY	1816	TTCTCTTTTGTATGTGTCTTCCAAAAGCGTACGTGTAGCTGATGTGTAAAGACGTGAA	1875
Db	1834	TTTTCTTTTGTATGTGTCTTCCAAAAGCGTACGTAGCTGATGTGTAAAGACGTGCAA	1893
OY	1876	GACAGGTGTGAAGTATTCAAAAAAATATTAAGATCTCATATATTTATTTGAATTTCTTGAA	1935
Db	1894	GATAGGTGTCAAAATATTTGAAAGAAATATTCGACGATCAAGATTTATGATTTCCCTTAAA	1953
OY	1936	AATAATACAAATTAATAATPACACACTGAGAAAGAAATTTGCTCTTGGCATACGATTTGC	1995
Db	1954	AAAAATCAATPAAAGACGAATCAGACGAAATTTGTGCTCATGSCATCCATACCTGC	2013
OY	1996	AATGATTTTACCTAATTTGTCCAGGTCTTACGAAAGAAATGT---TGTACAAATATC	2052
Db	2014	GATGATTTTATCCCAATGTCTCGATCTTAAAGAAAGAAATCTTTCGTCAAAATCTT	2073
OY	2053	AAGAAGCTCGTAGCCGTTCTTAAAGAAAGGCTTGGAAAGATGCTCTCAAGTATGAG	2112
Db	2074	AAAAATATTTGCGAACCATTTCTACAAAGAAAGGTTTACGAGATGCTCTTAAAGTATGAG	2133
OY	2113	CTTCAAGAAATTTGACTGATTAATCTPAAATGTGAACCTGCATTTGAAAGATATTGTACA	2172
Db	2134	CTTCAAGAAATTTAAGTAAATATACTAAATGTGAACCTGCATTTAAGAAAGATATTGTACA	2193
OY	2173	GTAGCGGGAACGTAAATATGTGCTCAATCAGTGGCTTATGCAAGCTAACACCAAGAT	2232
Db	2194	GTATTTGAAGACGTAAATTAATGCTCATACGAGTTTATGTAAAGATTAATCCGAAGT	2253
OY	2233	AACTGTGGAAG---AGTGAATGAGATGTGTAAGAAAGAACTGTGAGAAATTTAGTAAA	2289
Db	2254	AAAACTAAAGAGCCGCAATATPAAATATGTTAGAAAGAGCTTTGTCTPAAATTTAGTGGAA	2313
OY	2290	GAAGTGGAAGAACGTGCAAGCACTTACCAACAGATTTAGGACAAACCGGCACTGATCTA	2349
Db	2314	GAGGTGGAACGCAATGCAAGTATTTACCAACAGATTTAACAGAGCTGTGAAAAAAGCTTA	2373
OY	2350	AAAAAAGATTTAAGACATATGAGAACTTAAAGAACGTGCAAGGAAAGCAATGACAAAG	2409
Db	2374	AAAAAAGATTTAAGACATATGAGAACTTAAAGAAAGGCGCAAAAAAGCAATGACAAAG	2433
OY	2410	TCCAGCTGTGTTTGTCACTCACTTAAAGAAAGCAAGTATGTATCAAAAGTAAATAGC	2469
Db	2434	TCCAGCTGTGTTTATCACTTGTTAAGAAAAACAAAGTAAATACATGAAATAATATAGC	2493
OY	2470	AAAAACAAGATTAAGATGCGTTTCAACGCACTTCAAGATACACAAAAACATGTGAAA	2529
Db	2494	AAAAACAAGATTAAGATGCTTTTCAACGCACTTCAAGATACACAAAAATATGTGAAA	2553
OY	2530	ATACTACGAGGAGATTAAAGATGTATCCGTAACAGAAATTGAAGCTAAAGCATTTGAT	2589

D	b		2554	ATACTACGAAAGAGGGATTAAAGAGGCACTTGT7AACAAGATTGAAGCCAAAGCATTTGAT	2613
O	y		2590	TTGGCAGCAGAAGTAATTTGGAAAGATATGTAGATTGTAAGAAAGATGTAAATTGGAA	2649
D	b		2614	TTTGCCAGCAGAAAGCTTTGGAAAGATATGTAGCTTTGAAAGAAAATGTGAAGAAATTGACT	2673
O	y		2650	TCAGATTGCAGAAATTAAAGAGGATTGCCAAGACTTAGAAGAAAGTATGCGAAAAAGTTAAT	2709
D	b		2674	TCCGATTGGCGGGAATTTAAAGACGATTGGCATGGTATTAAAAGAAAGTGTGTGAAAAGTTGAG	2733
O	y		2710	AAGCCTTGTCCCAAATTGMAAGCCTCTGGAGGTTGAAGCCGCAGAAACAGTGCAGAAAGT	2769
D	b		2734	AAGACATGTCAACGATCTGAAAGCCTCTGGACGTTGAAGTGCATGAAATTGTCAACGAAAGC	2793
O	y		2770	ACAAAGCACTACCAACCAACCAACCAACCGCTTGGCGATCCGAAGGCAACGGAATGCAA	2829
D	b		2794	ACAACGACGACCACCAACCAACCAACGACCGTTACCGATCCGAAGGCAACGAAATGCAA	2853
O	y		2830	TTCTTACGAGAACAGACACATGGGTTACACAGACATGACACACACAAAGCACGTTACT	2889
D	b		2854	TTCTTACGAGAACAGATACATGGGTTACACAGACATGACACACACAAAGCACGTTACC	2913
O	y		2890	ATCACATCTACCATCACTCAAAAAATAACATTGACATCAACGAGGCGATGCAACCAAC	2949
D	b		2914	ATCACATCTACCATCACTCAAAAAATAACATTGACATCAACGAGGCGATGCAACCAAC	2973
O	y		2950	AAGTGTACGACAGAGAGGAAGATGATGTCAGAGACGTTGAAACCGAGTGAAGGGCTGAGG	3009
D	b		2974	AAGTGTACGACAGG-----GGATGAACAGAGAGACGTTGAAACCGAGTGAAGGGATTGAAG	3027
O	y		3010	ATGAGTGGGTGGAATGTGATGAGGGGGGTGATAGTACAAATGGTTATTTGGTTCAATGATT	3069
D	b		3028	ATGAGTGGGTGAGAGCGTGATGATGAGGGGGGTGATAGTACAAATGGTTATTTGGTTCAATGATT	3087
O	y		3070	TAG 3072	
D	b		3088	TAG 3090	

RESULT 5
AAZ94067
ID AAZ94067 standard; DNA; 3081 BP.
XX
AC AAZ94067;
XX
DT 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX
DE Pneumocystis carinii major surface glycoprotein gene HMSG32.
XX
KM Major surface glycoprotein; MSG; HMSG32; human; pneumonia; diagnosis; ds.
XX
OS Pneumocystis carinii; sp. f. hominis.

```

Key      Location/Qualifiers
FH       1..3030
FT       CDS                /tag= a
FT              /product= "HMSG32"
FT              /note= "a nucleic acid comprising residues 2836-3081 of
FT              this sequence is specifically claimed in Claim 28"
```

MO200009760-A1.
PD 24-FEB-2000.
PP 17-AUG-1999; 99WO-US018750.
PR 17-AUG-1998; 98US-0096805P.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
KI Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;

DR		WPI: 2000-206025/18.
DR	P-FSDS:	AAV79169.
XX		
PT	Detection of the presence of Pneumocystis carinii in specimens by	
PT	identification of major surface glycoprotein (MSG) gene sequences using	
PT	two or more oligonucleotide primers derived from human P. carinii MSG	
PT	protein encoding sequence.	
XX		
PS	Claim 27, Page 80-84; 110pp; English.	
XX		
CC	The present sequence is that of the novel Pneumocystis carinii sp. f.	
CC	hominis gene, HMSG32, which encodes a major surface glycoprotein (MSG,	
CC	see AAV79169). The gene was isolated by PCR amplification of DNA taken	
CC	from an autopsy lung sample of an HIV-infected patient with P. carinii	
CC	pneumonia. It is 1 of 7 novel, claimed genes (see AAV79063-69) of the	
CC	invention that encode human-P. carinii MSGs (see AAV79165-71). The MSGs	
CC	include a highly conserved C-terminal region of approximately 100 amino	
CC	acids; this region (also claimed) corresponds to residues 2836-3081 of	
CC	the present sequence. Direct detection or amplification of human-P.	
CC	carinii MSG-encoding genes, especially by PCR using primers directed at	
CC	the conserved region of the genes, provides a sensitive and specific	
CC	technique for the detection of P. carinii, and the diagnosis of P.	
CC	carinii pneumonia, especially in biological specimens (e.g. blood,	
CC	sputum) from immunocompromised patients such as those with HIV infection.	
CC	(Updated on 15-SEP-2003 to standardise OS field)	
XX		
SQ	Sequence 3081 BP; 1231 A; 431 C; 663 G; 756 T; 0 U; 0 Other;	
Query Match	52.9%; Score 1624.8; DB 3; Length 3081;	
Best Local Similarity	72.8%; Pred. No. 0;	
Matches 2275; Conservative	0; Mismatches 752; Indels 99; Gaps 10;	
OY	1 ATGCGCGGGCGGTCAAGCGGCAG-----GCACGACGGACACAGAAATGACTTGGAT	51
DB	1 ATGCGCGGGCGGTCAAGCGGCAGCGCTGCAAAAGCATCAGGGCGCTAATGATATGATGCT	60
OY	52 GAGGAACAATGTTTAGCTTAATTCTTAAGAAGATGAGCTAAGTGACGACGAATGCAA	111
DB	61 GAATAAATTCCTTTGGCTTTAATTGCA---GGAAAAAATTAATGATTAATGAATGCAA	117
OY	112 AAAAACCTTAATAAATTTGCGCAAGATTGCTGACGAAAACCTTAATATGAAACAAGTA	171
DB	118 AAAGAAATTTGAAAAATTTGTAAACATTTAAGGATGCAATTTAAACCAGAAAAAGTT	177
OY	172 CACAGAAAACTTAAGGTTTTTGGCAAGATGGAAGAGATACAAAATGCAAAAGACTG	231
DB	178 CACAAAAAAGCTTAAGAGTTTTTGTGAATAATAAAAAGAGAGTTCAAATATGCAAGAACTG	237
OY	232 AAAGCCAATATTGGAAGAAAAATGTACTCAATCAAGAGAAACTTTAAAGAACATTTAA	291
DB	238 AAAGAAAAAAGCTCACTCAAAAATGTACTGCATCAAGAGAAAACTTACAGAAAGCAATCAA	297
OY	292 AAAAANAATTCAGATTATTAACGAAATTAAGATTGCAAGAAAGAAATCAATGCCATATT	351
DB	298 AAAAANAATTCAGATTATTAACGAAATTAAGATTGCAAGAAAGAAATCAATGCCATATT	357
OY	352 TTGAGAGGAGATGATTTCAAAAAGAACTTAAGAGATTTGCAATATCTTTGAGAAATTAAGTC	411
DB	358 TTGAGAGGAGATGATTTCCAGCGGAACCTTAAGATATTTGCAATATCTTTGAGAAATTAAGTC	417
OY	412 TATCAAAAAGAAACGTGATAAAGTTGCCGAAAGAAATTTTAAAGACACTTTCGTAAGCGAT	471
DB	418 TATCAAAAAGAAACGTGATAAAGTTGCCGAAAGAAATCTTTTAAAGACAGTTTCGTAAGGT	477
OY	472 CTTAATGATGATGATCATATGTGAAAAAAAACCTTAAGAAGATTTGCCCTGTATGGGAGG	531
DB	478 CTAATCAATGAACCTACATGTGAAGAAAGCTCAAAAGGTTTCATAGAATTAAGTCAA	537
OY	532 GAAAGTATGATGATTAACAACCTTGTGCTGTAACAGAAAGAGATGATTAAGATTAATTTA	591
DB	538 GAAAGTATGATGATTAACAAGACCTTTGCTTTTATCAAAAAAAGAGTGCAGAAACATTTGTA	597
OY	592 ATTAAAAAAGATTAAGAAGTGGCTACTCTTAAAAAGATGTTTTCAGCAGACCTAG----	646

Db 598 TTGAAAAACAAAAATGTATCTCTTAAACAGATGTTTACGACGACTTGAGAG 657
Qy 647 -GAAGTTTAAAAAAGAACATGCTGTGATTAATCTGAAACAAATGCTATTTTAACTTGA 705
Db 658 AAGATGAGTTACGAGAAAAATGTTTACCACTGCTGAAACGATGCTATTTTATGAGAG 717
Qy 706 AATTGCGGAG-----ACGACGATTAAT 729
Db 718 AATTGTGAAGATATATCAAAATGTATAATATCATCCGAAGACTGTTATGAATATTGGCA 777
Qy 730 AAATGTATGTAATTGGAGAAAAATGCCAAGAACAAATTTGCTTATATGCAACCAAGA 789
Db 778 GTGTGTGATACATTGGCAGGTGAAATGTGAAAGAAATTAAGATTATTTATACATCCGGGA 837
Qy 790 CCCGATTTTATGATCAACTAGCCAGGCTACATATGACAGAGATATAGGGCTGGAGAG 849
Db 838 TCCGATTTCAATCAACTATAGCTAAAGCTTACTGACGAGAGACATGAGACTGGAGAG 897
Qy 850 TTTTATTAAGAAAGGTAGAGAGAGATGAGTTTATTTATGGAAGAAATCATCTAAGAGATCG 909
Db 898 CTTTATTAAGAAAGCGCAGAGAAAGATGTTTATTTGAGAAAGCTCTGTATAGAGATGCA 957
Qy 910 ACAGTTTGTGGCATTTGTTGATCCAAATTTAGTTAGTTTAAAAAAGAGACAAAGAG 969
Db 958 ACTGCTTACTGCGCTTTTGTATCAAAATCTAAGATCTTAAGATCAAGTGGTAAAG-- 1015
Qy 970 AAATCGAAGAAAGCCCTTCAAAAAAGCTCAAAAAATCTCATGACATGAGGCTTTAGAA 1029
Db 1016 -AATCGAAAAAGTTCTTAAAGATTAATCTTAAGAGTTAAAAAGTCATGAATTTTGGGA 1074
Qy 1030 AGTTTATGTAGAAAAATGTTTAAATGTAATGATGAAACGAAAAATGTGAAGATTGCA 1089
Db 1075 GATTTTGTAAATCAAAATGTAGCTGTCAAAATGAAATGAAAGTGAAGAGTTAAG 1134
Qy 1090 AATGATATTAACAAAATCTGCAAAATTTTCACTTCAAAAAGTCACTAATATATGTTT 1149
Db 1135 AAGAGGTATGACAAACAGTAAATCTTTTGAAGAAATTAAGAAATTAACACTCTCT 1194
Qy 1150 GATCCAAACAAAGAAATTAATGAATTTGTGATGGAAGGTTGCCAATTTCTTAGC 1209
Db 1195 -----GATCCGAGAAAGTCAATTCATGATTAAGTTTAACTTCTTAGT 1242
Qy 1210 AAGAGATTTGTGCAAAATTTGAGTCTTAATTTTCTAATTTTGAAGAAATGTCAGAT 1269
Db 1243 GACATGACTGCAAGAGTTTAAAGTCAAGTCTTTTATTTT---AAAAATCAAGCACT 1299
Qy 1270 GGAGAAATGCACTGTAATAATTAAGACAACTGTTACAAAAGAGACTTGATGACAG 1329
Db 1300 CTTGACAAAGAAATGTATATCTGAAGGCACTATTAAGAGAGGCTTGAAGCACAA 1359
Qy 1330 GCAATTAAGTCTGCAAGAAATATATGCGAAGAAATGTTACATGTTCAAAACAAAGCTG 1389
Db 1360 GCTATTAAGCACTGCAAGAAATATGTAAGGACTGTTCTATGCTTCAAGGCAAAAGATG 1419
Qy 1390 CTTGAAAAAGTTTCAACAAATTTAGTAAAGTATGAGAAACGTAAAAAGAAAAA 1449
Db 1420 TTTAAGAAATCTAGAAAAATTAATGAAAGATGTCGAACTTAAAAA----- 1470
Qy 1450 GGAAGTTTCTCAAGAGAAATTTATTTATTTCTGTGTGTAACGCCAGCAAAAGCACCCG 1509
Db 1471 -----ACAAAGGATGAGTTGTTTGTATGATTAATGATCCCTTAAACACTCGA 1521
Qy 1510 TTGCTTACATGATCTTGAATGAAAACTATCTTTTACGACAAACATGATGAAAAAG 1569
Db 1522 ATACTTGGAGCTGATATCCAAAGAGAGATCTTTTTCGAAAAAATTTGATTAAG 1581
Qy 1570 CGAGATTTCCGAGAGATTAATTAATGCAAGATTTGGGAGAAAGTCCAGATTTAGA 1629
Db 1582 CGAGACTTTCCAAAGAGAGATTTGCAAGATTTAGAAAGAAAGTGAAGCTTTAGGG 1641
Qy 1630 GAGGATTCAAAAGAAATTAATGAGGCAATGCTATCACTGAGAGCAATGCAATGCTTG 1689

Db 1642 AAGATTCAAATCAAGATTAAGTGGCATGTCATACGCTTAAAAACAAGTGTGATCGCTTG 1701
Qy 1690 GGAAGTACAGAAATTTTAAAGCAGTCTTTATTTGATGAAACAAAGATTAATTTGAAGAC 1749
Db 1702 GGAAGTACAGAAATTTTAAAGCAGTCTTTTACTATATGAACAAAGATTAATTTGAAGAC 1761
Qy 1750 CAAGAAAGTTGTGTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1809
Db 1762 CATGAAAGCTGTAAGAAATATTTAAAGAAAGAAATGTCATTAATGCTTGAAGGGGTGAT 1821
Qy 1810 GAGGTTTCTCTTTTGTATGTTGCTTCCAAAGCTACGTTGAGCTGATGTAAGAAC 1869
Db 1822 GATGTTTCTCTTTTGTATGTTGTTTACCAAAAGCTTACGTTGAGCTGATGTAAGAT 1881
Qy 1870 GTGAAGACAGTGTGAAGATTAATCAAAAAATTAATTAATTAATTAATTAATTAATTT 1929
Db 1882 GTGAAGACAGTGTGAAGATTAATCAAAAAATTAATTAATTAATTAATTAATTTCT 1941
Qy 1930 CTTGAAAAATTAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTT 1989
Db 1942 CTTAAAAATTAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTTCT 2001
Qy 1990 TATTGCAATGATTTTCACTTAATTTGTCCAGGCTTTACGAAAGAAATAGT---TGTACA 2046
Db 2002 TACTGATGATTTTATTTTACCAATTTGCTGATCTTAAAGAAAGAAATCTTTCTGCA 2061
Qy 2047 AAATTCAGAAAGCATGTCAGCCGTTCTATTAAGAAAGAAAGCTTGAAGATGCTTCAAA 2106
Db 2062 AATCTTAAAAAATTTGGAACCATTCCTACAAAAGAAAGTTTAAAGATGCTTTTAA 2121
Qy 2107 GTAGAGCTTCAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2166
Db 2122 GTAGAGCTTCAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTTGAAAGAT 2181
Qy 2167 TGTAAGTGGGAAACGTAAATTAATGCGTCAATCAATGAGTCTTATGCAAGTCAACAC 2226
Db 2182 TGCAATATTTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTCC 2241
Qy 2227 AAGATTAATCTGAAAGAGTATGAGATGCTGAAGAAAGAAAGCTCTGAGAAATTAAGT 2286
Db 2242 GAAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAGT 2301
Qy 2287 AAAGATGGAAGAACAGTCAAGACATTAACCAACAAATTAAGCAACCGCAGCTGAT 2346
Db 2302 GAAAGGTGGAACAGCAATGTAATGTTACACAGAAATTTGAGCAAGAGAAAGAAAC 2361
Qy 2347 CTTAAAAAAGTTTAAAGACATATGAGAACTTAAGAAAGCTGACAGAGAAAGCAATGAC 2406
Db 2362 CTTAAAGATGTTTGAACATTTGAACCTTTAAAGAAAGGCAAGAAACATGAT 2421
Qy 2407 AAGTCCAGTCTTGTGTTGTCATCTATTAAGAAAAAGAAAGTAAATGTAACAAAAGTAT 2466
Db 2422 AATTCATCTTGTGTTTATCATTTGTTAAGAAAGTGAATTAATATCATCGAAAAATG 2481
Qy 2467 AGCAAAAACAGATTAAGATTCGCTTCAACGGACTTCAAGATTAACCAACAAACATG 2526
Db 2482 AGCAAAAGCAAGATTAAGATTCGCTTCAACGGACTTCAAGATTAACCAACAAACATG 2541
Qy 2527 AAAATTAAGGAGGAGTTAAGATGATCCGTAACAGAAATTAAGATTAAGATTAAGAT 2586
Db 2542 AAAATTAAGGAGGAGTTAAGATGATCCGTAACAGAAATTAAGATTAAGATTAAGAT 2601
Qy 2587 GATTGGCAGAGAGATTAATTTGAAGATGATGATTTGAAGAAAGATTAATTAATTTG 2646
Db 2602 GATTGGTGAAGAGATTAATTTGAAGATGATGATTTGAAGAAAGATTAATTAATTTG 2661
Qy 2647 GAATCAGATTCAGATTAAGAGATTTGCAAGATTTGAAGAAAGATTAATTAATTTG 2706
Db 2662 GAATCAGATTCAGATTAAGAGATTTGCAAGATTTGAAGAAAGATTAATTAATTTG 2721
Qy 2707 AATAAGCTTTCGCAATCTGAAGCTCTGAGAGTGAAGCGGCAAGAAAGTGAAGAA 2766
Db 2722 CAAGAGATGTTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTG 2781

QY	2767	AGTCAACGCAACATCTACAAACAACAACAACCGTGGCATTCGAAAGGCAACGGAATGC	2826		
Db	2782	AGCAACAACGACGACCAACGACAAACAACGACCGTTACTGATCCGAGGCAACGAAATGC	2841		
QY	2827	AAATCCTTACAGACAAGACAACATGGGTTACAGACATCGACAACAACAGACGCT	2886		
Db	2842	AAATCTTTACAGACAAGATACATGATTCACAGACTTCGACACATACGACGCT	2901		
QY	2887	ACTATCAGATTTACCATCATCTCAAAAATAACATTGACATCAAGAGCGCATTCAAACA	2946		
Db	2902	ACCATCTCATTTACATCATCTCAAAAATAACATCTCATATCAACAGGCGTTGAAACA	2961		
QY	2947	ACCAAGGTACGACAGAGAGAGAAAGATATGTCAGAGAGCGTAAACCGAGTGAAGGCGCTG	3006		
Db	2962	ACCAAGGTACGACAGAGG-----GGATATGACAGAGACGTAAGACCGAGTGAAGGATTC	3015		
QY	3007	AGGATGAGTGGGTGGAATGTGATGAGGGGGGTGATAGTAGCAATGTTATTTCTTCATG	3066		
Db	3016	AAAGTAGATGGGTGAAAGTATGATGAGGGGGGTGATAGTAGCAATGTTATTTCTTCATG	3075		
QY	3067	ATTGAG 3072			
Db	3076	ATTGAG 3081			
RESULT 6					
AA294063	standard; DNA; 3042 BP.				
ID	AA294063				
XX	AA294063;				
XX	AC				
XX	AA294063;				
DT	15-SEP-2003	(revised)			
DT	05-JUN-2000	(first entry)			
XX					
DE	Pneumocystis carinii major surface glycoprotein gene HMSGp1.				
XX					
KW	Major surface glycoprotein; MSG; HMSGp1; human; pneumonia; diagnosis; ds.				
XX					
OS	Pneumocystis carinii; sp. f. hominis.				
XX					
XX	Key	Location/Qualifiers			
FT	CDS	1..3042			
FT		/*tag= a			
FT		/product= "HMSGp1"			
FT		/note= "a nucleic acid comprising residues 2894-3042 of			
FT		this sequence is specifically claimed in Claim 28"			
XX	MO200009760-A1.				
XX					
XX	24-FEB-2000.				
PD					
XX	17-AUG-1999;	99WO-US018750.			
PF					
XX	17-AUG-1998;	98US-0096805P.			
PR					
XX	(USSH)	US DEPT HEALTH & HUMAN SERVICES.			
PA					
XX					
PI	Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;				
XX					
DR	WPI; 2000-206025/18.				
DR	P-PSDB; AAY79165.				
XX					
PT	Detection of the presence of Pneumocystis carinii in specimens by				
PT	identification of major surface glycoprotein (MSG) gene sequences using				
PT	two or more oligonucleotide primers derived from human P. carinii MSG				
PT	protein encoding sequence.				
XX					
PS	Claim 27; Page 49-53; 110pp; English.				
CC	The present sequence is that of the novel Pneumocystis carinii sp. f.				
CC	hominis gene, HMSGp1, which encodes a major surface glycoprotein (MSG,				
CC	see AAY79165). The gene is 1 of 7 novel, claimed genes (see AA279063-69)				

CC of the invention that encode human-P. carinii MSGs (see AAY9165-71). The
CC MSGs include a highly conserved C-terminal region of approximately 100
CC amino acids; this region (also claimed) corresponds to residues 2894-3042
CC of the present sequence. Direct detection or amplification of human-P.
CC carinii MSG-encoding genes, especially by PCR using primers directed at
CC the conserved region of the genes, provides a sensitive and specific
CC technique for the detection of P. carinii, and the diagnosis of P.
CC carinii pneumonia, especially in biological specimens (e.g. blood,
CC sputum) from immunocompromised patients such as those with HIV infection.
CC (Updated on 15-SEP-2003 to standardise OS field)

XX
SQ Sequence 3042 BP, 1307 A, 381 C, 606 G, 748 T, 0 U, 0 Other;

Query Match 35.8%; Score 1098.6; DB 3; Length 3042;
Best Local Similarity 63.0%; Pred. No. 2.7e-217;
Matches 1954; Conservative 0; Mismatches 1054; Indels 93; Gaps 13

Oy	2	TGGGCGGCGGCGGTCAACGCGGACGACGAGGACACAGAAATGAGCATGATGAGGAA	58
Db	2	TGGGCGGCGGCGGTTAACGCGGACGTAAACAGAGCATCAGAGTATGATGAGGAAATGCG	61
Oy	59	ATGTTTATGCTTTAATTCTAAAGGAAGTGGACTAAGTACGAGGATGCAAAAAAAC	118
Db	62	GCTTTTGGCTTTATATCTAAAGAAAGATTCTAAGATGATTAATAATATGCGAAGAAAAAT	121
Oy	119	TAAAAAAATATGCGCAAGAAATTGACTGAAGAAACATAAATATATGAACAATACACAGA	178
Db	122	TAGAAAAACATGGAAGAAATTAAGTAGAAGCAATCTAATCTCCAGAACAAATGACATGAA	181
Oy	179	AACTTAAAGTTTTTGGCAAGATGAAAAAGCAGATACCAAAATGCAAGAACGTGAAGCCA	238
Db	182	AGTTAAAGATTTCTGTGATAGCAAAAAACGTATTAATAATATGTAAGAACTTAATAAAAA	241
Oy	239	ATATTTGAAAAAATGTACTACAAATCAAGAAAACTTAAGAACAATTAAAAAAA	298
Db	242	ATGTTGAAAAAATATGCGGTATTTTAAACAGAAATTAAGAAATTTGTGAAAAAAGGAAG	301
Oy	299	TTTCAGATTATACGGATTAAGGATTTGCAAGAAGATGAAACAACATGCTATTTTGGAG	358
Db	302	CTTCAATTTGAAAAATGATGATGTAACAAAAATGAACAACAGTCTGTTTTTTAGAAAG	361
Oy	359	GAGTATGTTTCAAAAGAACTTAAAGATGTTGCAATACTTTGAGAAATTAAGTATCAAA	418
Db	362	AAGCATGCTC---TGATCTTACAAAGAAATTGCAAGATTTAAGAAACAATGTTATCAGA	418
Oy	419	AGAAACGTGATAAAGTTGCGGAAGAAGTTCTTTAAGCACT--TGCTAGCCATCTTA	475
Db	419	ATAAGCGGATTAAGGTGACAAAGAAAGTTCTTTAAGATATATAAGAAAGAAATTTTA	478
Oy	476	ATGATCAGTCATATGTGAAAAAAAACCTTAAGAGATTTGGCCCTGTACGGGAGGAGAA	535
Db	479	AAGATTAATAATTCATGTGAATAATTAACGGAAGTACTGTCAAGAAATTAAGTCAATGCA	538
Oy	536	GTGATGAGTTAAACAACCTTGTGCTGCAACGAAAGAGACATGTATGAATATTTTAATTG	595
Db	539	GTGAGAAATGATGAATATATGTTTGAATCAAAAAAATACGTGTGATATCTTGTATTAAG	598
Oy	596	AAAAAGATTAAGAGTG-----CGTACTCTTAAACAGATGTTTCAGACGACATG	646
Db	599	AAAGCAACAAGAGTGGAATCTTTCAAAAATCTTAAACGGAATTTAAACAATTAAGG	658
Oy	647	GAACTTTTAAAAAAAAGAACATGTCTGAATTAACGCAACATGCAATTTTATACATTGGA	706
Db	659	AAGATGAACAATAAAAAAATATGCCCATATATATGAAGAAATGCATTTTTTATGATG	718
Oy	707	ATTGCGGAGACGAGATATTAATTAATGATATGGAATGGAGAAATGCCAAGACAA	766
Db	719	AAAGTTGTGGAAGACGATTCAGTGAAGTATGATGAAATGCAAAAAAATGTCAGAAGAAA	778
Oy	767	ATATTTGCTTATATGCGCACGAGACCCGATTTTGATCCAACTAAGCCAGAGGCTACATAG	826
Db	779	ATATTTACTTACACATTTATCATATTCAGGGTTTTGATCTATATGAACACAAATTTTCATTG	838

1133 CTAATTAATGCTCTTTTGTGATCCAAAGAAATATGAATTTGTGATGGAGGCT 1192
1124 GGATTTGAGAAATATTTCCCTCAGCTAATGTAAA---GGAATATTTGGTGGCATATGT 1180
1193 TGGCAACAATTTCTTTCGAAACGAAGTTGTGGCAAAATGGAGTCCATTTGTTTCTAATTTG 1252
1181 TGCATACATTTCTTGTGTGAAAAGAGGTGACGAACCTGTGTGGATTTGTTTATTTTGA 1240
1253 AAAAAAATGTCAGATGAGAGAAATGCAATGTAAAAATATAAGACAATGTTACAAA 1312
1241 AAAGCAAGCTCCTCT---GAAAGCCCTGCAATATCTTAAAGACACATGTTATAAA 1297
1313 GAGACTTGATGCACGGCAAAATAAGTCTGCAAGAAAATATGCGAAGATGTTACATG 1372
1298 AAGGCTTGAAGCAATGACCAATGAAAGATTACAGATTAAGTTAGGGGAAAATGCGAAG 1357
1373 GTTCAAAACAAGACGGCTGTGAAAAGTTTCAACAAGATTAGTAAATGATGTGAAAC 1432
1358 GTTCAAAATGAAACATGGCTTGAAACCTTCAAAAAACCTTGTAAGATTTGTGAAAAA 1417
1433 TGAATAAAGAAAAACAAGAGTTTCTCAACGATGAATTTATTTCTGTGTGACAGC 1492
1418 CGAAGAGAGA-----AAGTGAATTAATTTGTACTATGTATGAACC 1459
1493 CAGCAAAAGACGCCGCTTCTTACACATGATCTTGAATGAAAATCTATCTTTTACGAC 1552
1460 CAATTAATAACGGCTCTTACAGTGTCAACAGATTTGCGAATGAGGCAATGCTTTGCAAG 1519
1553 AAACAATGATCAAAAAGGAGATTTCCCGACAGATTAATAATGCAAGAAATGGGGAGAA 1612
1520 AGCATTTGAACGAAAAAGAGATTTTCCAAAGAAAAGATTTGAAGAAATTAAGAAAA 1579
1613 AGTCCCAAGATTTAGAGAGATTCAAAAAGAAATTCATGCGCATGTACATCACTGAGC 1672
1580 AATGGAAGGTCTTAGAAAAAGATTCAGAGAAATTAATGTGCACTGTATATGTTAAAA 1639
1673 AGCAATGCAATGCTTTGGGACTACAGAAATTTTAAAGCAGTTTATTTGATGACACA 1732
1640 AGCATTTGCAATCGCTGAAAGCATATGAGCACTTGAAGAGAGATGCTTAAAAAGAAATA 1699
1733 AAGATCTTTGAAAGCAAGAAAGTTGTGTAATTAATTAACCTTAAAAAAGTATAAAT 1792
1700 AAGGATATTTTAAAGATGAAATAGCTGCAAGAAAGAACTAAAGAAAGATGTAATAAT 1759
1793 GGTCTAGAAAGAGATGACCGTTTCTCTTTGTATGTGTCTTCCAAAACGCTACGTGTG 1852
1760 GGTTTAGAAAGAAATTAATAATTTTTCGGCTTGTCTGACTTGGAACTTGTGTGCA 1819
1853 AGCTGATGATTAAGACGTGAAAGACAGGTGTGAAGTATTCAAAAAATATATAAGCTT 1912
1820 AAAAGATCACTGAAATGTGTAATCTAAATGTAATATATTTGAAAAGCAATATGANAATA 1879
1913 CATATATTAATGAATTTCTGAAAATTAATACAAATTAATAACAACCTGAAAAAATAAT 1972
1880 TGAACGTTATTAAGTAAATAGCTAAATTAAGAGAAAATATGTGAATTT-----TT 1930
1973 GTCCCTCTTGCAATAGTATGCAATTAATTTTCACTTAATTTGTCAGAGTCTTACGAAG 2032
1931 GGGCTCCATATTTGTAAAAAGTACGAGCAAAATTTGAAAAAATCTTAAAAACGAGAAAAAG 1990
2033 AGAATAGTTTGAACAAAATCAAGAAAGCATGAGCCGTTCTATATAAAGAAAGCCCTTGG 2092
1991 ATGGCAATGCAAAAAAATCAATTAATAAGTCAAAATCAATCTTGAATAAAGAAAGCTTTAG 2050
2093 AAGATGCTCTCAAAAGTACAGCTTCAAGAAAAATGATGATTAATCTTAATGTAACCTG 2152
2051 AAAAATAAGTTGTGAAGAAATTTGAAGGTATGTTTATCAAAACGTAGAGAAATGTAACAATA 2110
2153 CATGAAAAAGATATTTGTACAGTAGCGGAAAGTAAATTAATGCTCAATCAAGTGGCTTAT 2212
2111 CACTTAATATATATCTGTACCAATTTGAAAAGGCAAGAAATGGCTTTGAAACTTTGT--- 2167

2213 GCAAAAGCTAACCAAGATTAATCTCTGAAAAGAGTGAATGAGATGCTTAAGAAAGAACTCT 2272
2168 -----GCAAAAGGAAAGAAAAC---CAAGATGACATTTAAAGTTAGAAAGAAACCTT 2218
2273 GTGAAATATTAAGTAAAGAAAGTGAAGAAACAGTSCAAAGCATTTACCAACAGAAATTAAGAC 2332
2219 GTGAAAAGCTATTAATAACGTTATAAAGAAATGCTCAAAATTTGAAGACAGACTTGAAG 2278
2333 AACCGGACGCTGATCTAAAAAAGATTTATAGACATATGAGAACTTAAAGAAAGCTGCAAG 2392
2279 AAGTAAAGAGGTCTTGAAGAAAAGAAAGAAAGATTAATAAATAATTAAGAAAGACAG 2338
2293 AGAAGCAATGAACAATCCAGTCTTTGTTTGTACATCATTTAAGAAAAACAAAGTAATG 2452
2339 AAAAAAGCATGGAAGATGCAAACTTATTTATGAGAGCGAAAGGACCTGATATATAT- 2397
2453 TATCAAAAAGTAAATAGCAAAAACAAGATTAAGATGCAATGCGTTTCAACGCACTTCAAGATA 2512
2398 -----AATATTAATGCTAGTAAATTAAGATCTCATCTGATACCTTAAGGAAGAAAGGCA 2452
2513 CCACAAAACATGTGAAAATATCTACGAGAGGAGTTAAGATGATCCGTAAACAGAAATTAG 2572
2453 CAACAGATTT-----AAACTGTAAGAAAGAAATGCAAAAGTGCATGTATACAGAAAAAG 2506
2573 AAGCTAAAGCATTTGATTTGGCAGCAGAAATTTGGAAGTATGTAGATTGTAAGAGAA 2632
2507 AATTAGCAGATTTGATTTGGTATGCAAGAGCATTTGATCTCTATCTGAATTTGAAAGAAA 2566
2633 GATGTAATTAATTTGGAATTCAGATTTGCAAGATTTAAGAGATTTGCAAGAACTTAAGAAAG 2692
2567 TATGTAATCATTTCTAGAGAAATTTGTGTCTTCAAAAAAGAGT---TGACTGTAGATC 2623
2693 TATGCAAAAAGATTAATTAAGGCTTTGTGCAATCTGAAGCCCTTGAAGGTGAAGCCGACG 2752
2624 CATGTAATAAAGATACAGGGAATATGTTCAACATTAAGGCCACTAAAGTGAACACACAG 2683
2753 AAACAGTGAAGAAATGACAAAGCAACCACTACAAACA-----ACAAACAACCGTTGCCG 2806
2684 AATATGTAATCTTAAAAACATTAACAATCAACCAACCAACCAACCACTACATTAAG 2743
2807 ATCCGAAGGCAAGGAATGCAAAATCCTTAACAGAACCAACAGACATGAGTTACAGACAT 2866
2744 ACGAAAGGCAACAGATCTGCCACTTTTACAGAACAGATACGTGGTCCAAAGACGT 2803
2867 CGACACACACAGACAGCTTATCTATCAATCTACATCAATCACAAAAAATTAACATTGACAT 2926
2804 CGACCCATATACGACATCCCAACCAACATCTACAGTCAAGTCAAGAAATAAGCTTGACCT 2863
2927 CAACGAGCCATGCAAAACCAACCAAGTGTACAGACAGAGAGAAATGATACAGAGACG 2986
2864 CGACAAAGACGGTGTAAAGCTTACGAAGGTACAGACAGAGAGAAATGAAACAGAGACG 2923
2987 TGAACGAGTGAAGGAGGCTGAGATGAGTGAATGTGATGAGGAGGAGTGAATGATGAG 3046
2924 TGAATCCGATGAAAGGTTGAGATGAGTGAATGAGTGAATGAGGAGGAGTGAATGAG 2983
3047 CAATGTTATTTGCTTCATGATT 3069
2984 CAATGACGATTTTCATTCATGATT 3006

RESULT 8
AA041223
ID AA041223 standard; DNA; 2110 BP.
XX
XX AA041223;
XX AC
XX 27-AUG-2003 (revised)
DT 17-DEC-2001 (revised)
DT 02-SEP-1993 (first entry)
XX
DE Clone PC3 encoding major surface gp of rat P. carini.
XX

KW Major surface glycoprotein; gp16; rat; *Pneumocystis carinii*; vaccine.
KW HIV; human immunodeficiency virus; diagnostic; PCR; ss.

OS **Pneumocystis carinii**

PN USN7958683-N.

PD 01-APR-1993

PF 09-OCT-1992; 92US-00958683.

PR 09-OCT-1992; 92US-00958683.

PA (USSH) US DEPT HEALTH & HUMAN SERVICE.

PI Kovacs JA, Angus CW, Powell F, Edman JC;

DR WPI; 1993-159487/19.

PT Major surface glyco-protein of *Pneumocystis carinii*, encoded by multi-copy gene family - used in a vaccine and diagnostic assay for *p. carinii* infection in AIDS patients.

PS Disclosure; Page 34; 50pp; English

Multiple clones were identified by immunoscreening a rat P. carinii cDNA library using rat serum generated against whole rat P. carinii. Clones reactive with polyclonal serum were evaluated to identify those encoding gp16. Three such clones (PC3, PC5 and PC14) were sequenced and contained gp16 encoding closely related but distinct proteins. Although none of the clones contained the complete coding sequence, overlapping regions allowed alignment of the three clones and generation of a putative composite sequence encoding a protein of ca. 122 kD. The gp16 of P. carinii can be used in a vaccine against infection in HIV-infected individuals and also as a diagnostic agent. Study of the expression of the protein should lead to understanding its role in the pathogenesis of P. carinii pneumonia and may lead to new strategies for control or prevention of infection. The DNA sequence may be used to make PCR primers for diagnostic use. NB. Due to a poorly reproduced sequence several unreadable bases are represented by N in the sequence below. See also AA041224-39. (Note: Revised entry submitted to correct the patent number format of US Government-owned NTIS applications to prevent clashes with ongoing US granted patent numbers. For further information please visit the Derwent web site at www.derwent.com/dmpi/updates/ntis_us.html.) (Updated on 27-AUG-2003 to correct OS field.)

Sequence 2110 BP; 745 A; 216 C; 480 G; 509 T; 0 U; 160 Other;

Query Match 9.48; Score 289.6; DB 2; Length 2110;

Matches 941; Conservative 0; Mismatches 789; Indels 78; Gaps 10;

Qy	23	GCACGACGGGACACAGAAATAGCACTTGATGAGAACATGTTTACGTTTAATTTCTAAAGGA	84
Db	351	GCACAAAGGAGCACAGATGACATTTAAAGGAGGAACCTTTTGCGTTTCATTCGCAAGAAC	410
Qy	85	GATGACCTAGTGAGCAGAGATGACAAAAAACTTAATAAATAATTTGCCAAGATTGACT	144
Db	411	GATATATGTTATGAGGATTAATGCAACCAAGACTCAGAAATATTTGTGAAGGTTGAG	470
Qy	145	GAAGCA-----AACTAATATTAAGACAGTACACAGAAAACCTTAAAGTTTTTCGGA	198
Db	471	GAAGCAGATGTTAAATTCATGTTATGATTAAGTTAAAGAACTTTGTGTGTGGTGAT	530
Qy	199	GATGAAAAAGCAGATCAAAATGCAAAAGACTGAAGCCATATTGAGAAAAATGTA	258
Db	531	GAAGCAAAACGAGATTAATAAATGCAAAAGCCTGAAGACCAAGTTGAAGTGAATTGAA	590
Qy	259	ACAACTAAAGAAAACTTTAAGAGCAATTTAAAAAAATTCAGATTTATACGATTA	318
Db	591	AAATTTGATGATGAACCTTCACAAAGCATTGAAGA-----CATTTAAAGATGAA	638
Qy	319	GATTGCAAAAGAGATATACAAACAAATGCTATTTTTTGAGGAGGTATGTCA--AAGA	375

Db	639	AATTGTGAAAAACATGAGAAAAATGTATCTTTTGAAGACACGGGTTATAGTAAAGAT	698
Oy	376	CTTAAAGATGATTCATTAACCTTTGAGAAATTAAGTGTATCAAAAAGAAACGTGATAAGTT	435
Db	699	ATTAAAGAAAGACGTGTCTCAAGTTGAGGGAAGGATGTTACAAATTTGAAGCGTTAAAGAGTG	758
Oy	436	GCGGAGAAGTCTTTTAAAGACACTTCGTGAGGATCTTAAATGATCATGTATGTGAA	495
Db	759	GCAGAGAGCTCCCTTTTGAAGGGCGCTCGAGGGGATGCTTAAAGATGAAAGCTTAATGTAA	818
Oy	496	AAAAAATTTAAAGATTTGCCCTGTACATGGGGAGGGAAAGTATGTATTAACAAACTTG	555
Db	819	GAAAAAGATGAAAACGTGTTGCTCCCAATTTAAGCCGAGAAAGTGAACGACTGATGTTTTC	878
Oy	556	TGCTCGAACAGAAAGACATGTAAAGATATTTTAAATGAAAAAGATTAAGATGCGGT	615
Db	879	TGCTTGATTTCCGATGGAACGTGTAAAGCGCTGAAAAAATAATCAAGAAAGATTTGCTG	938
Oy	616	ACTCTTAAACAGATGTTTCAAGACACTAGAAAGTTTAAAAAAAGAACATGTCTTGAA	675
Db	939	CCCTTAAAGAAAAAGCTTAAAGATGCGCAATT-----AAAGAAAAATGTCAGAA	989
Oy	676	TTATCTGACAAATGCTATTTTTCATTTGAAATTTGCGGAACGACGATATATTAATTAAT	735
Db	990	AGACTTGAAGAAATGCTCAATTTTTCAAAGAAAGCGTGACTTA-----AACAAATGT	1040
Oy	736	ATTGAATTTGGAGGAAATATGCGCAAGAACAAATATTTGCTATATATGACCACGAGCCGAT	795
Db	1041	GATAGAGATTTGAAGCAATGCAAGAAAAAGATTCACATATTAAGCCGCGGAATCTGAT	1100
Oy	796	TTTGATCCAACTAGCCGACAGGCTCAATAGCAGAGATATAGGCTGGAAGATTTTAT	855
Db	1101	TTTATGCTCTGTCAAGCCGAAGGCGGTGTTGTAGAGATATTTGGGTTTGATATGTGTAT	1160
Oy	856	AAGAAAGTAGAGAGATGAGTTTATTTTGAAGAAATCATCTAAGATGCGACAGCT	915
Db	1161	AAAAAGCTGAAAAAGAAAGATTTATTTGAAATCAGAGATGATCTTACCAAGGAAG	1220
Oy	916	TTGTGGCATTTGTGATCTCAAGATTTCAAGCTTAAAAAAAAGCAGCAAGAGAAATGC	975
Db	1221	TCAGTCAAAATTTCTGCAGAGATCTTGTCTACTGTGAGCAGAGATGGAATGATGCA	1280
Oy	976	GAAAGACCCCTTCAAAAAGCTGCAAAATCTCATGAATGAGGCTTTTGAAGTTTA	1035
Db	1281	GCGAAGAAATGCGGTAAAGGTTATGAAATGTGAACCTTAAATTTTGAATCTGAT	1340
Oy	1036	TGTAAAGAAAAATGTTTAAGTATGATGAAACGAAAAATGTGAAGATTGCCAAATGAT	1095
Db	1341	TTGATGAGATTATCAAAAGATGCTGATTAAGAAATTAATGCAAAAAAAGCTAGATG--	1398
Oy	1096	ATTAACAAACTTGCAAAATTTTCACTTCAAAAGTCACTAATATGCTCTTTTGATCA	1155
Db	1399	-TAAAGAAAGATGTACAAATCTCAAGTTAAATCTTATGTGAAGGGTGTGTACGAG	1457
Oy	1156	ACAAAGAAATAATGAA-----ATTGTTGATGCGAAGGGTTGCCAATTTCTTAGC	1209
Db	1458	TTTAAAGAAATTAATAATCAACTTTTATCGTGGGGAACGCTTCCACATTAATTAAG	1517
Oy	1210	AACGAAGATTGTGGAATTTGAGTCTTATTTGTTTCTTATTTGAAAAAAAATGTCCAGAT	1269
Db	1518	AAGGAGAGTGTGAGAACTTGAGTCCGAATGTTTCTATTTAGAAATGCGGTGAAGAT	1577
Oy	1270	GGAAATAAT-----GATGTAAAAATATAAGACCAATGTTTACAAAGAGACTTGAT	1323
Db	1578	AATAGATTTGTGAAGGCTGTCAAAATCTACGATCAGCGCTATTAATAAAGGGAACAAGAC	1637
Oy	1324	GCAAGGCAATAAAGTGTGCAAGAAATATGCGAGAGATGTTA---CATGGTTCAAC	1380
Db	1638	AGATGTTGAATTAAGTTCTTTCAAAAGGAATTAAGGAAAGCTTGTGTATGTAAAGTTT	1697
Oy	1381	AAAAGCTGGCTTGAAGATTTCAACAAGAAATTAGTAAAGTATGTGAGAACTGAAAAAA	1440

```
Db      1698 TATAGCGATCTTAAGATTGTAAAAATATGTGTGAGAAAAGTACAAACTTAAAAA 1757
Qy      1441 GAAAAAAGAGAAAGTTTCTCAAAACGATGAATATTATTCTGTGTGTACAGCCAGCAAAA 1500
Db      1758 GATTAATA-----GATACCTTTCAAAATGTCTTTATCTTAAAGAA 1796
Qy      1501 GCAGCCCGGTTCCTTACACATGATCTTGGATGAAAAACATCTTTTACGACAAACACTG 1560
Db      1797 CTATGTTATGGGCTTTCAAAATGATATTTTCTCCAAATCCAAAGATTAAAGTTCGCTTTTA 1856
Qy      1561 GATCAAAAGCGAGATTTTCCCGACAGATTAATAATTCAGAAATGGGGAGAAAGTCCAA 1620
Db      1857 GATGATCAGAGAGATTTTCCATTGTAAGAAAGATTTGCTTGAATTTGGAGAGAGATGTGAT 1916
Qy      1621 GATTTAGAGAGAGATTCAAAAGAAATTACATGCCATGTCAATACATGAGACGAACATGC 1680
Db      1917 CAATCTTAGAGATTCATTAATTAATTAAGAAAGTGTATTAATTAAGAAAGCGCTGT 1976
Qy      1681 AATCGCTTGGGACATACAGAAATTTTAAAGCAGTTTTATTTGATGAAACACAAAGATACT 1740
Db      1977 GAATATTTTTCAGCTTACAGAAAGATTTAGAAAAGTATTTTAAAGAAAGGATGATTCG 2036
Qy      1741 TTGAAGAACAAGAAAGTTGTATAAATCTATAAAGAAAGTGTATTAATGATCTAGA 1800
Db      2037 TTAATGATTCAGAGAAAGCTGTACAAAGGATTCATGAGAAATGTATTAATTAAG 2096
Qy      1801 AGAGAGAGA 1808
Db      2097 AGAGAGAAA 2104

RESULT 9
AAQ40201
ID      AAQ40201 standard; cDNA to mRNA; 2814 BP.
XX
AC      AAQ40201;
XX
DT      25-MAR-2003 (revised)
DT      09-AUG-1993 (first entry)
XX
DE      Sequence of jPB10 encoding the gp16 major surface antigen of rodent
XX      Pneumocystis carinii (PC).
XX
KM      Pneumocystis infection; diagnosis; AIDS; vaccine; therapy; sb.
XX
OS      Rattus rattus.
XX
FH      Key
FT      misc_feature
FT      Location/Qualifiers
FT      251..2814
FT      /*tag= b
FT      /note= "Identical to the homologous human Pc DNA"
FT      1617..2813
FT      /*tag= a
XX
PN      WO9307274-A1.
XX
PD      15-APR-1993.
XX
PF      30-SEP-1992; 92WO-US008328.
XX
PR      30-SEP-1991; 91US-00768166.
PR      18-OCT-1991; 91US-00761034.
XX
PA      (GEHO ) GEN HOSPITAL CORP.
XX
PI      Fishman JA;
XX
DR      WPI: 1993-134463/16.
DR      P-PSDB; AAR35082.
XX
PT      Isolated DNA encoding antigens of mammalian PNEUMOCYSTIS CARINII - used
PT      for prevention, treatment and diagnosis of PNEUMOCYSTIS CARINII
PT      infection.
```

```
XX      Claim 4; Fig 2; 74p; English.
PS
XX      The PC DNA, antigens and antibodies can be used for early diagnosis of PC
CC      infection, partic. in AIDS patients. The antigens and antibodies can also
CC      be used in vaccines against PC. The antigens, antibodies and antisense
CC      RNA can also be used in the treatment of PC infection. (Updated on 25-MAR
CC      -2003 to correct PN field.)
XX
SQ      Sequence 2814 BP; 929 A; 337 C; 657 G; 891 T; 0 U; 0 Other;
XX
Query Match      8.5%; Score 261.6; DB 2; Length 2814;
Best Local Similarity 51.3%; Pred. No. 4.1e-44;
Matches 879; Conservative 0; Mismatches 764; Indels 69; Gaps 9;
Qy      25 GCAGCAGGACACAGAAATAGCATTTGATGAGAGACATGTTTACCTTAAATCTAAAGAA 84
Db      1152 GTACAGATGACACAAGATGAGATTAAGAGAGAACCTTTGGCTTTCATTGTGAAAGAC 1211
Qy      85 GATGACCTAAGTGAAGCAGGAATGCAGAAAAAACTTAAAAATTTTGGCAAGATTGACT 144
Db      1212 AAACATGATGATGAGAAATGCAAAAAAGGCTCGAGAAATTTGTAAAGATGTAAAG 1271
Qy      145 GAGCAAAACTAAATATAGAACCAAGTACAGAAAACCTTAAAGTTTGGCAAGATGA 204
Db      1272 AAAGCAGATGAGAAATTT---CAGTGTGATGAGAAAGTTTAAAGACCTTTGTATGATAAA 1328
Qy      205 AAAGCAGATCAAAATGCAAAAGAACTGAAAGCCAATATTGAGAAAAATGTACTACATC 264
Db      1329 AAACGAGACGAAAATGCAAAAGAACTGAAAAAAAGTTAAAGATGAAATGGAACTTTT 1388
Qy      265 AAAGCAAACTTAAAGAGCAATTAATAAAAAAAATTCAGATTATTAAGGATTAAGATTGC 324
Db      1389 GATACGATCTTGAAGCATTCGGT-----AATGACATAGAAAGATGAAGAGATTTGT 1439
Qy      325 AAAGGAAATGAACAACATGCCCTATTTTGGAGGAGATGTTCAAAAGAACTTAAAGAT 384
Db      1440 AAAAAACATGAGAAAAATGTATACCTTTTAGAGAGACAGACCCAATATGCTTTAAGAG 1499
Qy      385 GATTGCATATCTTTAGAAATTAAGTGCATCAAAAGAAACGTGATTAAGTTGCGAGAGAA 444
Db      1500 AACTGTGTCAAGTTGAAGGAAAGATGTTAGCAATTTGAAGCCTTAAAGAGTGGCAGAGAG 1559
Qy      445 GTTCTTTTAAAGACATCTTGTAGCGATCTTAATGATCAATCATATGCTGAAAAAAACTT 504
Db      1560 CTCTTTTGAAGGCGCTCGAGGAGATGCTAAAGATGAAGCTTAATGTAAAGAAAGATG 1619
Qy      505 AAAGGATTTGCCCTGATGAGGAGGAGAAAGTATGATTAACAACTGTGTCTGAC 564
Db      1620 AAAACTGTTGCCCAATGTTAAGCCGAAAGATGACGAGTGAATTTTCTGCTTGAAT 1679
Qy      565 CAGAAAGACATGTAAAGATATTTTAATGAAAAAGATTAAGAAAGTGGCTACTCTTAA 624
Db      1680 TCGATGGAACGTGTAAAGCGCTGAAAAACAAATCAAGAAAGATTTGCTGCTTAA 1739
Qy      625 ACAGATGTTTACAGACATGAGAAATTTTAAAAAAGAAACATGCTTGAATTTACTCGA 684
Db      1740 GAAAAAGCTTAAAGATGCGGAAT-----AAAGAAAAATGTCAAGAAAGACTTAG 1790
Qy      685 CAATGCTATTTTTCATTGAAATTTGGCGAGACGACGATATTAATTAATGATTAAGATTG 744
Db      1791 AAATGTCAATTTTTCAAAAGAAAGCTGTACTGA-----ACAAATGTGATGAGAT 1841
Qy      745 GAGAGAAATGCCAAGAACAAATATTTGCTTATATGCAACCAAGACCCGATTTGATCA 804
Db      1842 ATGAAGCAATGCAAGAAAAAGATTCACATATAAGCGCGGAATGATTTTATGCTCT 1901
Qy      805 ACTAGGCCAGAGCTCAATATGACAGAGATATAGGCGCTGGAAGATTATTAAGAGGTA 864
Db      1902 GTCAAGCCGAAGGCGTGTGTTGAAGATATGAGGTTGATGATGATTAAGAAAGCCT 1961
Qy      865 GAGAGAGATGAGTTTATTTGAAAGAAATCATCTAAGAGATGGAACGCTTTGTGTGCA 924
```

Db	1962	GAAGAAAGGAATTTATTTATTTGAAAAATCAGAGTGGATCTACCAAGGAATCGAGTACA	2021
Qy	925	TTGTTGATCCAAAGTTCTAGTCTTAAAAAAAAGACGCAAGAAATGCGAAGAACCC	984
Db	2022	AAATTTTCTGCAAGATCTCTTGCTACTGTTCAGCAGAGATGGAATGATGCGAGAAAGAA	2081
Qy	985	CTTCAAAAAAGCTGCAGAAAAATCCCATGAAACATGAGCGCTTTAGAAAGTTATGTAAAGAA	1044
Db	2082	TGCGGTAAAGCGTTAGGAAAAATGTGAAACTTTAGTATTTTGAATCTGATTTGTGTGAG	2141
Qy	1045	AATGTTTAACTAAATGATGGAACGAAAAATGTGMAAGATTCGAAATGATATTAAACAA	1104
Db	2142	TTATGCAAAAGATGCGATTAAGAAAAATTAATATGCAAAAAAAGCTAGATG---TAAAGAA	2198
Qy	1105	ACTTGCAAAATTTTCACTTCGAAAGTCACTAATAATCGTCTTTTGTATCCAAACAAAGCA	1164
Db	2199	AGATGTACAAAACCTCAAGTTAAATCTTTATGTGAAGGGTTGTCTACGGAATTTAAAGAA	2258
Qy	1165	AATTAATGAA-----ATTGTTGANTGGGAAGGGTTGGCCACATTTCTTAGCAAGGAAT	1218
Db	2259	GATAAAAAATCACATCTTTTATTCGGGGGACAGCTTCCAAACATTTATTAACGAAGGAAG	2318
Qy	1219	TGTGCGAAATTTGAGTCCCTATTGTTTCTATTGTTTGAAAAAAATGTCAGATGAGAA--	1276
Db	2319	TGTGCAAGACTGTGAGTCCGAATGTTTCTATTTAGAAAAATGCGGTGAAGATTAATGAGTT	2378
Qy	1277	---ATGATGTAAAAAAATATAAGACCAATGTTACAAAAGAGACTTGATGACGGGCA	1332
Db	2379	GGTGAAGCGTGCAAAATCTACGATCAGCGTGTATTAATAAAGGACCAAGACAGATGTTG	2438
Qy	1333	AATTAAGTGTGCGCAAGAAATATGCGAGGAATGTT--ACATGTTCAACAAAAGCTGG	1389
Db	2439	AATTAAGTCTTTCAAAAGGAATTAAGGGAAGCTTGGTCATGTAAAGATTTTATAGCAT	2498
Qy	1390	CTTGAAGAAAGTTTCAACAGAATTAAGTAAGATGTGGAAGTGAAGAAAGAAAGAAACAA	1449
Db	2499	CCTAAAGATTTGTAATAATATGTGTAGAAACGTACAAAACTTAAAAAAGATTAAGAA	2558
Qy	1450	GGAAGTTTCTCAAAAGATGAATTAATTATCTGTGTGTACAGCCAGCAAAAGCAAGCCCGG	1509
Db	2559	TACCTTCAAAA-----TGCTTTTATCTTAAGAACTAATGTTAT	2597
Qy	1510	TTGCTTACACATGATCTTGGATGAAAACATCTTTTACGACAAACACTGGATCAAAAG	1568
Db	2598	GGGCTTTCMAATGATATTTTCTCCAAATCCAAAGATTAAGTTCCTTTTGAATGATCAG	2657
Qy	1570	CGAGATTTCCCGACAGATTAATAATTCGAAGAAATGGGGAGAAAGTGCACAGATTTAGCA	1629
Db	2658	AGAAATTTTCAATTTGAAAGAGATGTCTTGATATTTGGAGAGAGATGTGATCACTTACT	2717
Qy	1630	GAGATTCMAAAGAAATTAACATGCGCATGTCATACACTGAGCAGCAATGCAATCGCTTG	1689
Db	2718	AGTATTCATTATTGAAATTTAGAAAAGTGTAATTAACATTGAAAAGCGGTGGAATATTTT	2777
Qy	1690	GGGACTACAGAAATTTTAAAGCAGGTTTAAT	1721
Db	2778	GACGTTACAGAAAGATTTAGAAAAGTATTTT	2809
RESULT 10			
AAQ41226			
ID	AAQ41226	standard; DNA; 3521 BP.	
XX	AAQ41226;		
AC			
XX			
DT	27-AUG-2003	(revised)	
DT	17-DEC-2001	(revised)	
DT	02-SEP-1993	(first entry)	
XX			
XX			
DE	Clone GP3 encoding major surface gp of rat P. carini.		
XX			
KM	Major surface glycoprotein; gp116; rat; Pneumocystis carinii; vaccine;		
KM	HIV; human immunodeficiency virus; diagnostic; PCR; BB.		

XX	OS	Pneumocystis carinii.
XX	XX	USN7958683-N.
XX	XX	01-APR-1993.
XX	XX	09-OCT-1992; 92US-00958683.
XX	XX	09-OCT-1992; 92US-00958683.
XX	XX	(USSH) US DEPT HEALTH & HUMAN SERVICE.
XX	XX	Kovacs JA, Angus CW, Powell F, Edman JC;
XX	XX	WPI; 1993-159487/19.
XX	XX	Multiple clones were identified by immunoscreening a rat P. carinii cDNA
XX	XX	library using rat serum generated against whole rat P. carinii. Clones
XX	XX	reactive with polyclonal serum were evaluated to identify those encoding
XX	XX	gp116. Three such clones (PC3, PC5 and PC14) were sequenced and contained
XX	XX	ORFs encoding closely related but distinct proteins. Although none of the
XX	XX	clones contained the complete coding sequence, overlapping regions
XX	XX	allowed alignment of the three clones and generation of a putative
XX	XX	composite sequence encoding a protein of ca. 122 kD. PC5 was used to
XX	XX	screen a second cDNA library constructed in modified lambda ZAP, YCD11.
XX	XX	Four positive clones, GP3, GP22, GP46 and GP14 were sequenced and were
XX	XX	found to contain ORFs encoding proteins similar to the original gp116
XX	XX	clones. The gp116 of P. carinii can be used in a vaccine against
XX	XX	infection in HIV-infected individuals and also as a diagnostic agent.
XX	XX	Study of the expression of the protein should lead to understanding its
XX	XX	role in the pathogenesis of P. carinii pneumonia and may lead to new
XX	XX	strategies for control or prevention of infection. The DNA sequence may
XX	XX	be used to make PCR primers for diagnostic use. NB. Due to a poorly
XX	XX	reproduced sequence several unreadable bases are represented by N in the
XX	XX	sequence below. See also AA041223-39. (Note: Revised entry submitted to
XX	XX	correct the patent number format of US Government-owned NTIS applications
XX	XX	to prevent clashes with ongoing US granted patent numbers. For further
XX	XX	information please visit the Derwent web site at
XX	XX	www.derwent.com/dwpi/updates/ntis_us.html.) (Updated on 27-AUG-2003 to
XX	XX	correct OS field.)
XX	XX	Sequence 3521 BP; 1231 A; 413 C; 717 G; 740 T; 0 U; 420 Other;
XX	XX	Query Match 7.8%; Score 238.8; DB 2; Length 3521;
XX	XX	Best Local Similarity 48.6%; Pred. No. 2.36-39;
XX	XX	Matches 1061; Conservative 0; Mismatches 1052; Indels 71; Gaps 12;
XX	XX	285 AATTAAAAAAAATTGAGATTATACCGATTAAGATTGCCAAAGAAATGAACAACATG 344
XX	XX	421 ACTTCAAAAAGTATTGATATATATATATATATGAAAATTGGCAAAAATATGAAAGAAAATG 480
XX	XX	345 CCTATTTTGGAGGAGATATGTTCCAAAAGAACTTAAAGATGTAACCTTTGAGAAA 404
XX	XX	481 TATATCTTTTGAAGAAAGACGAGATTATATGATTATTAAGATTAATCTTATTCGAGTTGAGGA 540
XX	XX	405 TAACTGCTATCAAAAAGAAACGTGATTAAGTTGCGAAGAAAGTTCTTTTAAAGACACTTCG 464
XX	XX	541 AGGATGTTTAACAATTGAAGCGTGAAAAGGTTGGGCAAGAGAGAGTTCTTGAAGGGCGTCGG 600
XX	XX	465 TAGCGATCTTAATGATCAGTCATATGTGAAAAAAAATTAAAGACATTTGGCCTTCAT 524
XX	XX	601 AGGGAATGCTTAAAGAAAGAGCTTAATATTAAGGAAAGATGAATACCTGTTGGCCAGTGT 660
XX	XX	525 GGGGAGGGAAGAGTAGATTTAAACAACCTTGTCCTGAACCAAGAAAGACATTAAGAA 584
XX	XX	661 GAGCGGAGAAAGCAGCAATTAATGATGCTTTTGGCTTGATTTCTGTAACATGTGGAGA 720

QY 585 TATTTAATGAAAAAGATGAAGAGTCGGTACTCTTAAAAACAGATGTTTCAGACGACT 644
 Db 721 TCTGAAAAAAATTTGGGTACTGTTTGGAGCCTTTAAAAAAAAGAGCTT-----AA 771
 QY 645 AGGAAGTTTAAAAAAGAAACATGCTTGAATTAATCTGACAAACATCTATTTTACATTGG 704
 Db 772 AGATTAACAAATTTAGCGAAAGTGTCATGAAGACCTTAAGAAATGTCATTTTTCAGGAGA 831
 QY 705 AAATTCGCGAGACGACGATATATTAATTAATGATTAATGGGAGGAAATGCCAAGACA 764
 Db 832 AGCGTGTATGATG-----CGAAATGCAAGAAAGTTTGAAGGCAATGCAAGGAGAA 882
 QY 765 AAATTTGCTTATATGCAACGACGACCCGATTTTGATCCAACTAGAGCCAGAGCTTACAT 824
 Db 883 AAATTTATATTAAGAGCGCCAGATCTGATCTGTCTGTCAAGCCGAGGCGCTCTT 942
 QY 825 AGCAGAGATATATGGGCTGGAAAGATTTTATTAAGAAAGTGAAGAGATGAGATGTTTAT 884
 Db 943 GTTGAGAAAGTATTTGGGTTGGATGATGTATTAACCGGAAAAACATGGGATTTATAT 1002
 QY 885 TGGAAA-----GAATCATCTTAAGAGATGCGACAGCTTTGTTGGCATTTGATCCAGA 938
 Db 1003 TGGAAAAATCAGAGAGTGATCTTACCAAGAGATGACGTAACAATTTCTGCAAGATCTT 1062
 QY 939 TTCTAGCTTTAAAAAAAAGACGACAAAGAAATGCGAAGAGCCCTTCAAAAAGCTG 998
 Db 1063 GCTACTGTTGAGCAGATGAGATGAAGAGAACCAAGATAAAAAGTGCACACTAAAGCGTT 1122
 QY 999 CAAAAATCCTCATGAACATGAGGCTTTAGAAAGTTTATGTAAAGAAATGTTTTAAGTA 1058
 Db 1123 AGAAAAATGTATGTCCTCTAAGTATTTGAATTAATAATTTGAAAAAGTTATGTAAAGATG 1182
 QY 1059 TGATGGAACGAAAAATGTGAAGATTCGCAAAATGATATTAACAACCTTGCAAAATTTT 1118
 Db 1183 AAACAAAAACGAAAAATGCAAAAAATATAGATGTAA--AAGAAAGATGTACAAATCTC 1240
 QY 1119 CACTTCAAAAGTCACTAATATATGCTTTTGTATGCAACAAAAGGAAATATGAAT--T 1176
 Db 1241 AAATTAATACTTTATCTGAAGAGATGTCACGGAATATGATGATCAAGAAATCAAGATCT 1300
 QY 1177 GTTGAATGGGAAGGTTGCCAATTTCTTACCAAGAAATGTCGAAATTTGGAGTCC 1236
 Db 1301 TTATCTGTGGGACACGTGCCAATTTTATTAAGAGAGTGTGCAAGACTTGAAGTCG 1360
 QY 1237 TATGTTTCTATTTTGAAGAAAAATGTCAG-----ATGAGAAATATGCAATGTAAAT 1290
 Db 1361 GAATGTTTCTATTTGAAGAAAGCGGTGAAGATATATATTTGATTAAGCGTGCAGAAAT 1420
 QY 1291 ATTAAGACAACATGTTTACAAAAGAGACTTGATGCAACGGCAATTAAGTCTGCAAGAA 1350
 Db 1421 GCAGAGACGCGTCTATTAAGAAAGGACAGAGATGTTGAATTAAGTCTTCAAAAG 1480
 QY 1351 AATATGCGAGGAATGTT--ACATGTTCAACAAAGCTGCTTGAAGAAAGTTTCAACAA 1407
 Db 1481 GAATTAAGAGGAAAGCTTGCTATGTAAAGTTTATAGCATCTTAAGATTTGTAAGAAA 1540
 QY 1408 GAATTAAGTAAGTATGAGAACTGAAGAAAAAGAAAACAAGAAAGTTTCTCAAGCAT 1467
 Db 1541 TATGTGCTGTAAGAAACTGTACAAAACTTGATTA----- 1573
 QY 1468 GAATTAATTAATTTCTGTGTATCAGCCAGCAAAAAGAGCCCGTTCCTTACATGATCTT 1527
 Db 1574 AAATATCTTCCACAGATGCTTATCTTAAAGAACTATGTTATAGGCTTTCAAAATGATAT 1633
 QY 1528 CGAAATGAAGAACTATCTTTTATAGCAACAACATGATCAAAAGCGAGATTTCCCGACAAT 1587
 Db 1634 TTTCTTCAATCCAAAGATTAAGTGTGCTTTTGTGATGATCAAAAGGATTTTCCATTAAA 1693
 QY 1588 AAAAATTCGAGAAATGGGAGAAAGTGCACAAAGTTTATAGAGAGATTTCAAAAGAAAT 1647
 Db 1654 AAGGATTTGTGTGAGTGAAGAGAGTGTGATGAACCTTAGTAGATTTCAATTATGAAAT 1753

QY 1648 ACATGCCATGTCTATACATCTGAGACAGCATGCAATGCTTTGGGACTACAGAAATTTTA 1707
 Db 1754 TTTAGAAAAAGTATATACATTGAAAAAGACGTGTGAATATCTTTAGAGTTTCAGAGGATTT 1813
 QY 1708 AAGCAGGTTTATTTGATGAACACAAAGATATCTTTGAAAAGCAAGAAAGTGTGTAAA 1767
 Db 1814 AGAAATGATTTTATGAAAAAAAAGATGATTCGTTAAATGACTCGAGTAACTGTACAAAG 1873
 QY 1768 TACCTAAAGAAAAGTGTAAATTAATGCTTGAAGAGAGATGAACGTTTCTCTTTGTA 1827
 Db 1874 GCATTTGATGAGAAATGCAATCAATTAATATGAGAGAGAAAGAAATTCATTTAGTGTTC 1933
 QY 1828 TGTGCTTCCAAAACGCTAGCTGTGACGTATGTTAAAGACGTAAGACAGGTGTGA 1887
 Db 1934 TGTCTTTTACCGAAGAAACATGTATATATGTTATTTCCATACAGTCAAGATTTGAGT 1993
 QY 1888 GTATTTCAAAAAAATATTAAGCTTCATATATATTAATTTGAAATTTTGAATAATATACAAT 1947
 Db 1994 AGTTTAAAGTCACATCAAGATGAATAAATTTTGAATAAATTTGGAGAAATTTTAA 2053
 QY 1948 AAAATTAACAACCTGAA--AGAAATGTCCCTCTTGCGATACGTATTCGAATGATTTT 2005
 Db 2054 AAACCAATATAAATGAAGCCTTGCTTGAAGAACTTGCACAAATGCGGCGCATTTGT 2113
 QY 2006 CACTTAATTTGTCAGGCTTTACGAAGAGATATGTTTACAAAAATCAAGAAAGATCGTG 2065
 Db 2114 CACCACTTAATGAGAAATTTGCCGA--TGACTTAAAAAAAAGAAAGATGCAATGGCAA 2172
 QY 2066 AGCGTCTTATTAAGAAAGGCTTTGGAAGATGCTCTCAAGTGAAGCTTCAAGGAAAT 2125
 Db 2173 TGATCATTAATCTGCGAGACATCTCAAGAAATATGCAATTTTGAAGATGTTAAAGAAAGC 2232
 QY 2126 TGACTGATTA--ATCTAAATGTGAACCTGCTATGAAAAGATTTGTACATGACGGGAA 2182
 Db 2233 AGAGAGAGAGCTGATGATCTGTTGAAGGACGATTTAAAGATGATTAATGTAAAGAAAGC 2292
 QY 2183 ACGTAAATTAATGCTCATCTAGTGTCTTTATGCAAGCTAACACCAAGATTAATCTGGA 2242
 Db 2293 ATTAAGAAAGGCTTGACCTGATGTTGAAGAG--AATGAAGCAATTCGTAATGTGTA 2351
 QY 2243 AGATGATGAGAGATGCTAGAAAGAAAGCACTGTGGAATTAATGTAAGAAAGTGAAGAAC 2302
 Db 2352 AATGTGATGATTAATCAAGAAAGATTTTGCAAAAAATTTAGTGTATTAAGTAAAAAGA 2411
 QY 2303 AGTCAAAAGCTTACCAACAGATTTAGACAAACCGGACGTGATCTAAAAAAGATTATA 2362
 Db 2412 GATGCCCTACTTTTAAAGACGAACTGGAGATGCGAAAAAAGAGTTGACAAAGATGAAGA 2471
 QY 2363 AGACATATGAGAACTTAAGAAACGTGCAAGAGAAAGCAATGAACAAGTCCAGTCTGTTT 2422
 Db 2472 ATGAGTACGATGATCTCAAAAAGCGGCAAAAAATCTTACGAGGCAAGCTTAAGTTATTGC 2531
 QY 2423 TGTCACTCATTAAGAAAAAGGAA 2446
 Db 2532 TATCAAGACTTAAGCAAACTGTAA 2555

RESULT 11
 AAQ41230
 ID AAQ41230 standard; DNA; 3521 BP.
 XX
 AC AAQ41230;
 XX
 AC 27-AUG-2003 (revised)
 XX
 DT 17-DEC-2001 (revised)
 XX
 DT 02-SEP-1993 (first entry)
 XX
 DE Gene encoding major surface gp of rat P. carini.
 XX
 KW Major surface glycoprotein; gp16; rat; Pneumocystis carinii; vaccine;
 XX
 KW HIV; human immunodeficiency virus; diagnostic; PCR; ss.
 OS Pneumocystis carinii.

	Key	Location/Qualifiers
FH	mlec_feature	1..722 /*tag= b
FT	CDS	/note= "fragment determined by PCR" 146..3412 /*tag= a
FT	mlec_feature	626..3521 /*tag= c
FT		/note= "fragment from original GP3 clone"
PN	USN7958683-N.	
XX		
PD	01-APR-1993.	
XX		
PP	09-OCT-1992;	92US-00958683.
XX		
PR	09-OCT-1992;	92US-00958683.
PA	(USSH) US DEPT HEALTH & HUMAN SERVICE.	
XI	Kovacs JA, Angus CW, Powell F, Edman JC;	
PI	WPI: 1993-159487/19.	
XX	P-PDB: AAR36706.	
DR		
Major surface glyco-protein of Pneumocystis carinii, encoded by multi-copy gene family - used in a vaccine and diagnostic assay for p. carinii infection in AIDS patients. Disclousre: Page 41; Sopp; English.		
Multiple clones were identified by immunoscreening a rat P. carinii cDNA library using rat serum generated against whole rat P. carinii. Clones reactive with polyclonal serum were evaluated to identify those encoding gp16. Three such clones (PC3, PC5 and PC14) were sequenced and contained ORFs encoding closely related but distinct proteins. Although none of the clones contained the complete coding sequence, overlapping regions allowed alignment of the three clones and generation of a putative composite sequence encoding a protein of ca. 122 kD. PCR was utilised to determine the full sequence of the gene. The 5' end of the message was identified by anchored PCR using primer JK58, which spanned the putative start codon of the composite protein. The intervening region was determined by reverse transcription followed by PCR using primers spanning the 5' end to base 722 in GP3. A single clone was identified that had an identical sequence to the first 76 bases of GP3. The CC complete, composite cDNAs contained an ORF encoding a protein of ca. 122 kD. The gp16 of P. carinii can be used in a vaccine against infection in HIV-infected individuals and also as a diagnostic agent. NB. Due to a poorly reproduced sequence several unreadable bases are represented by N in the sequence below. See also AAQ41223-39. (Note: Revised entry submitted to correct the patent number format of US Government-owned NTIS applications to prevent clashes with ongoing US granted patent numbers. For further information please visit the Dervent web site at www.dervent.com/dwpl/updates/nctls-us.html.) (Updated on 27-AUG-2003 to correct OS field.) Sequence 3521 BP; 1238 A; 385 C; 695 G; 725 T; 0 U; 478 Other; Query Match 6.9%; Score 211; DB 2; Length 3521; Bee Local Similarity 41.7%; Pred. No. 1.3e-33; Matches 1030; Conservative 0; Mismatches 1357; Indels 82; Gaps 14		
OY	1 ATGGCGGCGGCAGTGAACGCCAAGCAG---CAGGACAACAGAATAGCATTTATGAGAA	57
Dd	146 ATGGCAACGCGGTTAAGAGCCAGCACACTGACAGACGACAGATTGATGAGAA	205
OY	58 CATGTTTTAGCTTTAATTCATAAGAGAACATGACCTTAGTGACCAGAAATGCAAAAAAA	117
Dd	206 CAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAGAAACAAAATGCAAAGAA	265
OY	118 CTAAAAAAATATTCACAGATTGACTGAAACAAAACCTAAATATAGAACAACTAC-CAG	176

D	b		266	CTCGAGAAATATTGTAAGAAGTTGAAGAGACAGATATAAATGTAGCAATGTGGATGAT	325
O	y		177	AAAACTTAAAGTTTTTGCGAAGATGGA AAAACGAGATACA AATTCGA AAACACTGGAAGC	236
D	b		326	AA	385
O	y		237	CAATATTGAGAAAATATGTA CTACAATCA AAGAAAACCTTA AAGAACATTTAAAAAAA	296
D	b		386	AAAAGTTGAAGTAGAATTTAA AAGTTTTGAGAGGAACCTT CAAAAGTATTGAATTAAT	445
O	y		297	AATTCA GATTATA CGATTAAGATTGCAAAGAAATGA CAACAA TGCTATTTTTGA	356
D	b		446	AA-----ANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATTACTTTTGA	492
O	y		357	GCGAGTATGTTCAAAGAACTTAAAGTATGTTGCAATCTTTGAGAAATTAAGTCTATCA	416
D	b		493	AGACACGAATTATGATGTTATTAA GGATTAAC TGTATGAGTTGAGGGAAGATGTTACAA	552
O	y		417	AAAGAAACGATTAAGATGCGGAAGAA GTTCTTTTAAGAGCACTTGAGCATCTTAA	476
D	b		553	ATTGACGCGTGA AA	612
O	y		477	TGATCAGTCATATGTGAAAAAA ACTTAAAGATTTGCCCTGTCA TGGGAGGGAAG	536
D	b		613	AGAGAAGCTTAATGTAAAGAAAAGATGA ATCTGTTGCCAGTGTGAGCGAGAAAG	672
O	y		537	TGATGAGTTAACAACTTGTGTCTGA ACAGAAAAGACATGTBA GAATTTTAAATGA	596
D	b		673	CGACGAATTGATGCTTTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCGA AAAAAA	732
O	y		597	AAAGATTAAGAGTGC GGCTACTCTTAAAGATGTTCCAGACACTAGGAAGTTTAA	656
D	b		733	ATTGGGTA CTGTTTGCA GCCCTTTAAAAAAGACTTAAAGATPACGAATTAG-----	785
O	y		657	AAAGAAACATGCTTGAATTACTCGAACATGCTATTTTACATTGGAATTCGAGAGA	716
D	b		786	--CGAAAAGTGCATGAAGA CTTGAAGATGTATTTTANNNNNNNNNNNNNNTGA	843
O	y		717	CGACGATTAATTAATGTATTGA ATTGGAGGAAAAATGCCAAGACAAATATTTGCTTA	776
D	b		844	TG-----CGAAATGCAAGAA GTTTGAGAGCAATGCAAGGGA AAAAAATATATATA	894
O	y		777	TATGCCACAGAGCCGATTTTGA TCACATAGSCCAGAGGCTACATPATCAGAGATAT	836
D	b		895	TAAAGCCCGAAACTATCTTACTCTGTCAAGCCAGAGGCGCTCTTGTTNNNNNNNN	954
O	y		837	AGGCGTGAAGAGTTTATATAAGAGTAGAGAGATGAGTTTATTTTGA AAAAATC-	895
D	b		955	NNNNNGATGATGTGTATTA AAAACGCGGAAAAACATGGGATTTATTTGGAAAAATCAG	1014
O	y		896	----ATCTA GAGATCGACA GCTTTGTTGGCATTTGTTATCCAA GATTCAGTCTTAA	950
D	b		1015	AGTGAATCTACAGAGAAAGTCAGGTACA AATTTCTGCAAGATCTCTTNTCANNNNNN	1074
O	y		951	AAAAAAGACACAAGAGAAATCCGAAGACCCTTCAAAAAAAGCTGC AAAATCTCTCA	1010
D	b		1075	NNNNNTGAGATTAAGAA GMAACGAGTAA AAGCTGA CTAAAGCGTTAGAAAAATGTGA	1134
O	y		1011	TGAACATGAGGCTTTAGAAAGTTATATGA AAAAAATGTTTTAAGTATGATGGAACGA	1070
D	b		1135	TGCCTTAAGATTGTGAATCTGAATTTGGAAA GTTATGTBAAGATGGA AACAAAANN	1194
O	y		1071	AAAAATGTAAGAA TTGCAAAATGATATTA CAAACTTGCAAA TTTTCACTTCAAAAGT	1130
D	b		1195	NNNNNNCAAAAAATATTA GATGTA--AAGAAAGATGTACA AATTC AATTTAAACTT	1252
O	y		1131	CACTAATATATGCTCTTTTGA TCACAAAAGGAAATATAGAAATTTGTGATGGGAAG	1190
D	b		1253	TATCTGAAGATTTGTCTACGA TAATGAATCMAGAACTCAGATCTTTATGCTGNNN	1312
O	y		1191	GTTGCAACATTTCTTGAACA CGAAGATGTGCGA -AATTGGAAGCTTATGTTCTATT	1249
D	b		1313	NNNNNNNACTTTTTTATTA BAAGGAGAGTGTCA GAACCTTGAGTCCGAATGTTTTCTATT	1372

```

QY 1250 TTGAAAAAAATGTCAGATGAGAAAAATGCAATGTAATAATTAAGACCAATGTTACA 1309
    |||||
Db 1373 TAGAAAAAGCGCTGTAAAGAT-----AATAATATGATTAAGCGTCAAAAATGACAGA 1425
QY 1310 AAAAGAGACTTGATCCAGCGGCAATTAAGTGTCCAGAAAAATATGAGAGAAATGTTAC 1369
    |||||
Db 1426 GCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1481
QY 1370 ATGTTTCAACAAAGCTGTGCTTGAAGTTTCAACAAGAAATTAAGTAAGATGTGAGA 1429
    |||||
Db 1482 --AATTGAAGGAAAGACTTGTGCATGTAAGATTTATGCGACTCTAAGTTGTAABAAA 1539
QY 1430 AACTGAAAAAAGAAAAAAGAAAGATTCTCAACAGATGAATTAATTATCTGTGTAC 1489
    |||||
Db 1540 TATGTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1595
QY 1490 AGCCAGCAAAAGACCCGGTTCCTTACACATGATCTCGAATGAATCTATCTTTTAC 1549
    |||||
Db 1596 ATCTTAAAGAACTATGTTATGGGCTTCAAGATATATTTCTTCAATCCAAAGATTAA 1655
QY 1550 GACAAACAATCGATCAAAAGCAGATTCCGACAGATTAATAATGCAAGAAATGGGGA 1609
    |||||
Db 1656 GTGGGCTTTTGGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1715
QY 1610 GAAAGTCCAAAGATTAGAGAGATTCAAAGAAATTAATGATGCGATGTCATACATGG 1669
    |||||
Db 1716 AGAAGTGATGATGATCTAGTAGTATCTATTAATTAAGAAAGNGTAAACANNNN 1775
QY 1670 AGCAGCAATGCAATCGCTTGGGACTAGAAATTTTAAGAGGTTTATTTGATGAAC 1729
    |||||
Db 1776 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1835
QY 1730 ACAAAGATACTTTGAAGACCAAGAAAGTGTGTAAAAATACCTTAAGAAAAAGTAAATA 1789
    |||||
Db 1836 AGAATGATTCGTATATGATCTAGATTAATCTGACAAAGCAATGATGAGAAATCCATC 1895
QY 1790 AATGCTTGAAGAGAGATGACCGTTCTCTTTTGTATGTGTCTTCCAAAACGCTACGT 1849
    |||||
Db 1896 AATTATATAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1955
QY 1850 GTGAGCTATGCTTAAGACGTAAGACAGGTGGAAGTATTCAAAAAATAATTAAG 1909
    |||||
Db 1956 GTAGTTATATGATATTCATACATCAAGATTTGATGATTTAAAGTCAACATCAAGA 2015
QY 1910 CTTCAATATATTTATGAAATTTCTTGAATAATATACAAATTAATAACAACATGAAAG-- 1967
    |||||
Db 2016 ATGAAAAAATTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2075
QY 1968 -AAATGTCCTCTTGCGATACGATTCGATATGATTTTACCTTAATTGTCAGGCTTTA 2026
    |||||
Db 2076 TTGTTTGAAGAACTCTGACACAACATGGGCGCGACATTTGACCACTTATGAGAAATG 2135
QY 2027 C-----GAAAGAGATAGTGTACAAAAATCAAGAGATGATGAGCGGCTTCTATA 2077
    |||||
Db 2136 CCGGATGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2195
QY 2078 AAAAGAAAGCCTTGGAGATGCTCTCAAGTAGAGCTTCAAGSAAATTTGACTGATAAT 2137
    |||||
Db 2196 CAAGAAAAATGCAATTAACCTTTGAAAAGTTGAATTAAGAGAGAGACTGATCATCTG 2255
QY 2138 CTTAAATGTGAACCTGATTTGAAAAATATTGTACATGACGGGAAACGTAATAATGCGT 2197
    |||||
Db 2256 TTGAAGGATTTTAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2315
QY 2198 CAATAGTGGCTTATGCAAAAGCTAACCAAGATTAACCTGGAAGATGATGAGATG 2257
    |||||
Db 2316 TGGAAA-----AGAAATGAAGCAATTAACCTCTGATGTAAATGTGTATGAATA 2366
QY 2258 CTAGAAAGAACTGTGTGAAATTTAGTGAAGAAAGATGGAAGAAAGTGAAGCAATTAAC 2317
    |||||
Db 2367 CCAAGAGAAATGTGTGCAAAAAATTTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2426

```

```

QY 2318 CACAGAAATTTAGCAACCCGACGCTGATCTTAAAAAAGATTATTAAGACATATGAGAAC 2377
    |||||
Db 2427 AAGACGAATCTGGAGAAATCCGAAAAAGATTGAACAAAGATGAAGATGACATGATC 2486
QY 2378 TTAAGAAACGTGACAGAGAAAGCAATGAACAGTCCAGTCTGTTTGTCACTCATTAAGA 2437
    |||||
Db 2487 TCAAAAAGGCGCGAGAAAAATCTACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2546
QY 2438 AAAACGAAA 2446
    |||||
Db 2547 AAAGCTTAA 2555
    |||||

RESULT 12
AAQ40202
ID AAQ40202 standard; DNA; 1448 BP.
XX
AC AAQ40202;
XX
DT 25-MAR-2003 (revised)
DT 09-AUG-1993 (first entry)
XX
DE Sequence of JS7-2A2U encoding the gp116 major surface antigen of rodent
DE Pneumocystis carinii (PC).
XX
KW Pneumocystis infection; diagnosis; AIDS; vaccine; therapy; ss.
XX
OS Rattus rattus.
XX
PN W0307274-A1.
XX
PD 15-APR-1993.
XX
PF 30-SEP-1992; 92WO-US008328.
XX
PR 30-SEP-1991; 91US-00768166.
XX
PR 18-OCT-1991; 91US-00781034.
XX
PA (GENO ) GEN HOSPITAL CORP.
XX
PI Fishman JA;
XX
DR WPI; 1993-134463/16.
XX
XX
XX Isolated DNA encoding antigens of mammalian PNEUMOCYSTIS CARINII - used
XX for prevention, treatment and diagnosis of PNEUMOCYSTIS CARINII
XX infection.
XX
PS Claim 4; Fig 3; 74p; English.
XX
CC The PC DNA, antigens and antibodies can be used for early diagnosis of PC
CC infection, partic. in AIDS patients. The antigens and antibodies can also
CC be used in vaccines against PC. The antigens, antibodies and antisense
CC RNA can also be used in the treatment of PC infection. (Updated on 25-MAR
CC -2003 to correct PN field.)
XX
SQ Sequence 1448 BP; 537 A; 174 C; 377 G; 360 T; 0 U; 0 Other;
XX

Query Match 6.2%; Score 192; DB 2; Length 1448;
Best Local Similarity 54.1%; Pred. No. 8.6e-30;
Matches 565; Conservative 0; Mismatches 440; Indels 39; Gaps 7;
QY 35 CACAGAAATGACATTTGATGAGAACATGTTTATGCTTAATTTCTAAAGAGATGACTAA 94
    |||||
Db 432 CACAGATGAGATTTAAGAGAACACCTTTGGCTTTCATTTGTAAGACAAACATGATG 491
    |||||
QY 95 GTGAGCAGAAATGCAAAAAAATTAATTTGCCAAGATGCTGAAGCAAAAC 154
    |||||
Db 492 ATGAGAAATGATGCAAAAAAGCGTCGACGAATTTGTAAGAGATTGAAGAAAGCAGATG 551
    |||||
QY 155 TAAATATGACACAGTACAGAAAACTTAAGGTTTGGCAGAGATGAAAAAGCAGATA 214
    |||||
Db 552 AGAATTT---CAGTGTGAATAGAAAGTTTAAGACCTTTGTGATGATATAAAAAAGAGAC 608
    |||||

```

QY 215 CAAATGCAAGAACTGAAAGCCATATTGAGAAAAATGTAATCAATCAAGAAAC 274
 DB 609 AAAAAATCAAGAACTGAAAGAAATCTGGGAGTGAATTTGGAACTTTGATACGATC 668
 QY 275 TTAAGAAGCAATTAAGAAAAATTCAGATTATACCGATTAAGATTGCAAGAGATG 334
 DB 669 TTGAAGCATCGGT-----AGATGACATGAAAGATGAAGAGATTGTTAAAAACATG 719
 QY 335 AACACAAATGCTTATTTTGAAGGAGTATCTCAAAAGAACTTAAGATGATTGCATA 394
 DB 720 AAGAAAAATGTATCTTTTAGAGAGACGACCCAAATAGCTTTAGAGAACTGTGTCA 779
 QY 395 CTTTGAAGAAATTAAGTCTATCAAAAGAAAGCTGATTAAGTTGCGGAGAAAGTTCTTTAA 454
 DB 780 AGTTGAGGAGAAAGATGTTACGAAATTTGAAGCTTAAGAGTGGCAGAGAGAGTCTTTTGA 839
 QY 455 GAGCACTTCGTAGCGATCTTAATGATCAGTCATATGTGAAAAAACTTAAGAGATTT 514
 DB 840 GGGGCTCGGAAAGAGAGCTAAAGAGATTAATGTAAGCAGAGATGAAAAAGGTTT 899
 QY 515 GCCTGTCAATGGGAGGAAAGTATGATGATTAACAACTTGTGTCTGACCAAGAAAGA 574
 DB 900 GCCCAGGTAAAGCCGAGAAAGCAGCAATGATGTTTGTGCTTGTGATTCGATGAA 959
 QY 575 CATGTAGAAATATTTTAAAGAAAGTAAAGTGGGTACTCTTAAACAGATGTTT 634
 DB 960 CGTCTGAG---CGCTGAAAAAAATCAGAAAGTTTCCAGCTTTTAAAGAAAGCTTA 1016
 QY 635 CAGCAGCACTAGAAAGTTTAAAAAGAAACATGCTTGAATTAATCGAACATGCTATT 694
 DB 1017 AAGATG-----GCGAATTAAGAAAGAAATGTCAATGAAAGCTTGAGAAATGTCAAT 1067
 QY 695 TTTTCATTTGAAATTTGGGAGACGACATATTAATTAATGATTTGAAATGGAGAAAT 754
 DB 1068 TTTTAC-----GAGAAAGCGTGTGATTAACAAATGTGATGAGATTAAGATCAAT 1118
 QY 755 GCCAAGAAACAAATATTTGCTATATGCGACAGGACCGATTTTGAATCAACTAGGCGAG 814
 DB 1119 GCGGAAAAAAAGAAATCATAATTAAG---CGCCAAATTCGATTTCAATGCTCTCAAGCGA 1177
 QY 815 AGGCTACATATGACAGAGATTAAGGCTGAAAGATTTTATTAAGAGTAAAGAGAGATG 874
 DB 1178 AGAGCTGTTGTTGAGAAAGTATGGGTTGATGATGATTAAGAAAGCTGAAAGAG 1237
 QY 875 GAGTTTTTATTTGAAAGATATCTTAAGAGATGCGACGCTTTTGGCATTGTTATCC 934
 DB 1238 GAATTTATTTGAAATTCAGAGAGTGTCTTACCAAGAGAGTCAAGAAATTTCTGC 1297
 QY 935 AAGATT-----CTAGCTTAAAAAAAAGACGAAAGAAAGAAAGCGACCTTCA 989
 DB 1298 AAGATCTCTTCTAGTCTTTCGACAGAGTGAATGATTAAGAGTCAAGGAGAAATGCG 1357
 QY 990 AAAAAAGTCAAAAAATCTCATGAACATGAGGCTTTAGAAAGTTTATGTAAGAAAAATG 1049
 DB 1358 AAAAAAGGTTAAAAAATGTAAGCTTCTAGTATTTGATTAATCTGATTTGATGAGTTATG 1417
 QY 1050 TTTTAAGTAAATGATGAACGAAAA 1073
 DB 1418 CAAAGATGCTGATTAACAAAAAA 1441

RESULT 13

AAQ41224
ID AAQ41224 standard; DNA; 1454 BP.

XX AAQ41224/
 AC
 XX 27-AUG-2003 (revised)
 DT 17-DEC-2001 (revised)
 DT 02-SEP-1993 (first entry)
 XX Clone PCS encoding major surface gp of rat P. carinii.

XX Major surface glycoprotein; gp116; rat; Pneumocystis carinii; vaccine;
 KM HIV; human immunodeficiency virus; diagnostic; PCR; ss.
 OS Pneumocystis carinii.
 XX USN7958683-N.
 PN
 XX
 PD 01-APR-1993.
 PF 09-OCT-1992; 92US-00958683.
 PR 09-OCT-1992; 92US-00958683.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 XX Kovace JA, Angus CW, Powell F, Edman JC;
 PI WPI; 1993-159487/19.
 DR
 XX
 PT Major surface glyco-protein of Pneumocystis carinii, encoded by multi-
 PT copy gene family - used in a vaccine and diagnostic assay for P. carinii
 PT infection in AIDS patients.
 PS Disclosure; Page 35; 50pp; English.
 XX Multiple clones were identified by immunoscreening a rat P. carinii cDNA
 CC library using rat serum generated against whole rat P. carinii. Clones
 CC reactive with polyclonal serum were evaluated to identify those encoding
 CC gp116. Three such clones (PC3, PC5 and PC14) were sequenced and contained
 CC ORFs encoding closely related but distinct proteins. Although none of the
 CC clones contained the complete coding sequence, overlapping regions
 CC allowed alignment of the three clones and generation of a putative
 CC composite sequence encoding a protein of ca. 122 kD. The gp116 of P.
 CC carinii can be used in a vaccine against infection in HIV-infected
 CC individuals and also as a diagnostic agent. Study of the expression of
 CC the protein should lead to understanding its role in the pathogenesis of
 CC P. carinii pneumonia and may lead to new strategies for control or
 CC prevention of infection. The DNA sequence may be used to make PCR primers
 CC for diagnostic use. See also AAQ41223-39. (Note: Revised entry submitted
 CC to correct the patent number format of US Government-owned NTIS
 CC applications to prevent clashes with ongoing US granted patent numbers.
 CC For further information please visit the Derwent web site at
 CC www.derwent.com/dwpi/updates/ntis_us.html.) (Updated on 27-AUG-2003 to
 CC correct OS field.)
 XX
 SQ Sequence 1454 BP; 545 A; 164 C; 360 G; 385 T; 0 U; 0 Other;
 Query Match 6.2%; Score 191.2; DB 2; Length 1454;
 Best Local Similarity 49.9%; Pred. No. 1.3e-29;
 Matches 719; Conservative 0; Mismatches 663; Indels 60; Gaps 7;
 QY 307 ATAAAGATTAAGATTGCAAGAAAGATGAACAACATGCTTATTTTGAAGGAGTATGT 366
 DB 55 ATAAAGATTAAGATTGTAAGAAACATGAAGAAAAATGTTACTTTTGAAGGAAAGCAGAC 114
 QY 367 TCAAAAGAACTTAAGATGATGCAATCTTGAAGAAATAAGCTCTCAAAAGAAAGCT 426
 DB 115 CCAATATGCTTAAGAGAAAGTGTGCAAGTGAAGGAGAGATGTAAGATTAAGACGT 174
 QY 427 GATTAAGTTGCGAAGAAAGTTCTTTTAAGAGCACTTGATGCGATCTTAATGATCAGTC 486
 DB 175 GAAAGAGTGCAGAGAGAGCTCTTTTAAGGCACTCGAGAGGATGCTAAAGAAAGATGCT 234
 QY 487 ATATGAAAAAAACTTAAGAGATTTGCCCTGTCAATGAGGAGGAAAGTATGATGATTA 546
 DB 235 AAATGTAAGAAAGAAAGATTAATCTGTTGCCCAAGTTTAAGCGAAGAAACGCAAGATTG 294
 QY 547 ACAAACTTGTGTGCAAGCAAGAAAGACATGTAAGAAATTTTAATGAAAAAGATAG 606
 DB 295 ATGACTTTTGTCTTGAATCGAGATGAGACGTGTGAGAGCTGAAAAACAAATTTGGCGGAA 354
 QY 607 AAGTGCCTACTCTTAAGACAGATGTTTCAGCAGCACTAGAAAGTTTAAAAAAGAAACA 666

```

Db      355 GTTGGCAACCTTTGAAAACGAAATGGAATGAGAAA-----AGCTCAAAAAAG 402
Qy      667 TGTCTGAATTAAGTGAACAATGCTATTTTATGATGGAATTTGGGAGACGACGATATA 726
Db      403 TGTCTGAAAAGCTTGAGAAATGTCAATTTTTCAGAAAGAGCGTGTG-----TAAAT 453
Qy      727 ATTAATGTTATGAAATGGAGAGAAATGCGCAAGAACAAATATGCTTATATGCGACCA 786
Db      454 ACAAAATGTAAGAGATTAAGACGAAATGCGAGAAAAGAAATGCAATATTAAGCGCG 513
Qy      787 GGACCCGATTTTGAATCAACTAGGCGAGGCTAATAAGCAGAGATATAGGGCTGGAA 846
Db      514 GAATCTGATTTTATGTCGGTCAAGCGGAAAGCGGTGTTGAGAGATTTGGGTTGAA 573
Qy      847 GAGTTTATTAAGAAAGTGAAGAGATGGAGTTTATTTGGAAGAAATCATCTAAGAGAT 906
Db      574 GATGTGTTAATAAAGCGGAAAACATGCGATTTATTTGGAAAATCAGAGATGATCTA 633
Qy      907 GCGACAGCTTGTGGCAATGTTGATCCAAAGATTCATCTTAAAAAAAAGACGACAAA 966
Db      634 CCAAGGAAGTCAGTACAAAATTTCTGCAAGATCTTGTGCTAGCTTGAGCAGAGATGAG 693
Qy      967 GAGAAATGCGAAGAACCCCTTCAAAAAGCTGCAAAAATCCTCATGAACATGAGCTTTA 1026
Db      694 AATGATGAGGAGAAATGCGGTAAAGCGTTAGAGAAATGATGCTTCAAGATTTTG 753
Qy      1027 GAAAGTTATGTAAGAAAATGTTAAGTATGATGAGACGAAAATATGTCAGAAATTTG 1086
Db      754 GATCATTAATTTGAAAGATTAATGCAATGATGAGAAAAGAAAAGCAAAATGCAAAATTA 813
Qy      1087 CAAAATG---ATATTAACAAAATTGCAAAATTTTCACTTCAAAAGTCACTAATATATGCT 1143
Db      814 CTAGATGTAAATGTAAAGAAAGATGTACAAAATCTAAATCTTAATGTGAAAGGG 873
Qy      1144 CTTTTT-----GATCCAAACAAGAAATATATGAATTTGTGATGGGAGGCTTGCA 1197
Db      874 TTGCTTACAAAATTTGAAAAGAGCTGAAAATCAGATCTTTTATGCGTGGGAGACGCTTCA 933
Qy      1198 ACATTTCTTACGACGAAAGATTTGCGAAATTTGAGATCCTATTTGTTTCTAATTTGAAAA 1257
Db      934 ACATTTATTTACGAAAGGAGAGTGTGCAAGAACTTGATCGGATGTTTCTAATTTAABAAAT 993
Qy      1258 AAATGTCCAGATG-----AGAAAATGATGTAAAAATATATAGACCAATGTTACAA 1311
Db      994 GCGGTAAAGATATATAGATTTGATGATGATGATGCAAAATGCAAGACGCGCTATATAA 1053
Qy      1312 AGAGGACTTGATGCAACGGGCAATTAAGTGTGCGAAGAAAATATGCGGAGAA---TGTTA 1368
Db      1054 AAGGACACAGACAGATGTTGAATAGTTCTTTCAAAAGAAATTAAGGGGAAATCTTGCT 1113
Qy      1369 CATGCTTCAACAAAAGCTGCTGTGAAAAGTTTCAACAAGATTAAGTAAAGTATGTGAG 1428
Db      1114 CTGTGAAGTTTATATAGCATCTGTAAGATGCAAAAATCTGTGTGAGAACTGTACA 1173
Qy      1429 AAATGAAAAAGAAAACAAAGAAATTTCTCAAAAGATTAATTAATTTCTGTGTGTA 1488
Db      1174 AAATCTTAAAGAGATATGATATACCTTTCAAAA-----TGTCTT 1212
Qy      1489 CAGCCAGCAAAAGCAGCCCGTGTGCTTACATGATCTTGAATGAAATCTATCTTTTAA 1548
Db      1213 TATCTTAAAGATTAATGTTATGCGCTTCAAAATGATTTTCTTCAATCAACAAAGATTA 1272
Qy      1549 CGACAAACAATGATCAAAAGCGAATTTCCGACAGATTAATAAATTTGCAAGAAATTTGGGG 1608
Db      1273 AGTTGCTTTTGGATGATCAAAAGGATTTTCCATTATGAAAAGATTTGCTTGAATTTGGGTG 1332
Qy      1609 AGAAAGTGCAGATTTAGAGAGAGATTCAAAAGAAATTAATGATGCGCATGCTATACACTG 1668
Db      1333 GAGAAGTGTGAATCTTATGATGATTCATTAATGAAATTTGAAAAGATGATTAACATTTG 1392
Qy      1669 GAGCAGCAATGCAATCGTTGGGACTACAGAAATTTTAAACAGAGTTTATTTGATGAA 1728

```

```

Db      1393 AAAAGACGCTGTGAATCTTTAAGCTTACAGAGGATTTAGAAAAATTTTAAAAAA 1452
Qy      1729 CA 1730
Db      1453 AA 1454

RESULT 14
AAZ94070
ID AAZ94070 standard; DNA; 249 BP.
XX
XX AAZ94070;
XX AC
XX AC 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
XX
XX DE P. carinii major surface glycoprotein gene HMSGp2 conserved region.
XX
XX KM Major surface glycoprotein; MSG; HMSGp2; human; pneumonia; diagnosis; ds.
XX
XX OS Pneumocystis carinii; sp. f. hominis.
XX
XX PN MO200009760-A1.
XX
XX PD 24-FEB-2000.
XX
XX PF 17-AUG-1999; 99WO-US018750.
XX
XX PR 17-AUG-1998; 98US-0096805P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX PA Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q,
XX PI MPI; 2000-206025/18.
XX
XX DR P-PSDB; AAY79172.
XX
XX PT Detection of the presence of Pneumocystis carinii in specimens by
XX PT identification of major surface glycoprotein (MSG) gene sequences using
XX PT two or more oligonucleotide primers derived from human P. carinii MSG
XX PT protein encoding sequence.
XX
XX PS Claim 27; Page 103; 110pp; English.
XX
XX CC The present sequence is that of DNA encoding the conserved C-terminal
XX CC portion (see AAY79172) of the novel Pneumocystis carinii sp. f. hominis
XX CC major surface glycoprotein (MSG), HMSGp2. The invention provides novel
XX CC genes (see also AAZ99063-69) that encode human-P. carinii MSGs (see also
XX CC AAY79165-71). The MSGs include a highly conserved C-terminal region of
XX CC approximately 100 amino acids. Direct detection or amplification of human
XX CC -P. carinii MSG-encoding genes, especially by PCR using primers directed
XX CC at the conserved region of the genes, provides a sensitive and specific
XX CC technique for the detection of P. carinii, and the diagnosis of P.
XX CC carinii pneumonia, especially in biological specimens (e.g. blood,
XX CC sputum) from immunocompromised patients such as those with HIV infection.
XX CC (Updated on 15-SEP-2003 to standardise OS field)
XX
XX SQ Sequence 249 BP; 73 A; 49 C; 77 G; 50 T; 0 U; 0 Other;

Query Match 6 0%; Score 185; DB 3; Length 249;
Best Local Similarity 83.9%; Pred. No. 1.5e-28;
Matches 209; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy      2821 GAATGCAAAATCCTTAACAGACAAACAGACATGGGTTACACAGATCGACACACAAGC 2880
Db      1 GAGTGCCAAATCTCTGACAGACAGACACGCTGGGTCAAAAGAGCTGACCAATACTAGC 60
Qy      2881 AGCTTACTATACATCTTACATCAATCAAAATTAACATTTGACATCAAGAGCGATGC 2940
Db      61 ACTTCTACGATAGTCCACAGTCAATCGAAGATTAACACTCACCTCAACGAGCGGTGT 120
Qy      2941 AAACCAACCAAGTGTACACAGAGAGAAAGATGATGAGAGAGCGTAAACCGAGTGA 3000

```

Db 121 AAGCTTAGAAGTGTACACACAGAGAGAGAAGATATATGACAGAGAGGTGAACCCAGATGAA 180

Qy 3001 GGGCTGAGAGATGAGTGGCTGGAATGTGATGAGGGGGGGTGTATGATGACATGCTATTTCG 3066

Db 181 GGGCTGAGAGATGAGTGGCTGGAATGTGATGAGGGGGGGTGTATGATGACATGATTTTCA 240

Qy 3061 TTCTATGATT 3069

Db 241 TTCTATGATT 249

RESULT 15

AAQ41225

AAQ41225 standard; DNA; 2190 BP.

XX AAQ41225;

XX AC

XX 27-AUG-2003 (revised)

DT 17-DEC-2001 (revised)

DT 02-SEP-1993 (first entry)

XX

XX C1one PC14 encoding major surface gp of rat P. carinii.

XX KM Major surface glycoprotein; gp116; rat; Pneumocystis carinii; vaccine;

XX KM HIV; human immunodeficiency virus; diagnostic; PCR; ss.

XX

XX Pneumocystis carinii.

XX OS

XX USN7958683-N.

XX PN

XX 01-APR-1993.

XX PD

XX 09-OCT-1992; 92US-00958683.

XX PF

XX 09-OCT-1992; 92US-00958683.

XX PR

XX (USSH) US DEPT HEALTH & HUMAN SERVICE.

XX PA

XX Kovacs JA, Angus CW, Powell F, Edman JC;

PI WPI; 1993-159487/19.

XX

XX

PT Major surface glyco-protein of Pneumocystis carinii, encoded by multi-

PT copy gene family - used in a vaccine and diagnostic assay for P. carinii

PT infection in AIDS patients.

XX

XX

PS Disclosure; Page 36; 50pp; English.

XX

CC Multiple clones were identified by immunoscreening a rat P. carinii cDNA

CC library using rat serum generated against whole rat P. carinii. Clones

CC reactive with polyclonal serum were evaluated to identify those encoding

CC gp116. Three such clones (PC3, PC5 and PC14) were sequenced and contained

CC ORFs encoding closely related but distinct proteins. Although none of the

CC clones contained the complete coding sequence, overlapping regions

CC allowed alignment of the three clones and generation of a putative

CC composite sequence encoding a protein of ca. 122 kD. The gp116 of P.

CC carinii can be used in a vaccine against infection in HIV-infected

CC individuals and also as a diagnostic agent. Study of the expression of

CC the protein should lead to understanding its role in the pathogenesis of

CC P. carinii pneumonia and may lead to new strategies for control or

CC prevention of infection. The DNA sequence may be used to make PCR primers

CC for diagnostic use. See also AAQ41223-39. (Note: Revised entry submitted

CC to correct the patent number format of US Government-owned NTIS

CC applications to prevent clashes with ongoing US granted patent numbers.

CC For further information please visit the Derwent web site at

CC www.derwent.com/dwpi/updates/ntis_us.html.) (Updated on 27-AUG-2003 to

CC correct OS field.)

XX

XX

SQ Sequence 2190 BP; 874 A; 320 C; 473 G; 523 T; 0 U; 0 Other;

Query March 4 8%; Score 146.2; DB 2; Length 2190;

Beet Local Similarity 48.3%; Pred. No. 2.9e-20;

Matches 622; Conservative 0; Mismatches 623; Indels 42; Gaps 6;

OY	1172	AAATGTTGATGCGGAAGGTTGCCAATTTCTTACCAAGAAATTTGGCGAAATTTGG	1231
Db	8	AACTTTATGCTGGGGACAGCTTCCAACCTTTTATTAAGAGAGATGTGCAGAACTTG	67
OY	1232	AGTCTTATTTGTTCTATTTTGGAAAAAAATGTCAGATGGA--GAAATGCATGTAAA	1288
Db	68	AGTCGGAATGCTTCTATTATTTGAAAAATCGGTGTACGAATTAAGATTGTGAAGCATGTCAA	127
OY	1289	ATATTAAGCAACATGTTTACAAAAGAGACTTGTATGCACGGCGAAATTAAGTGTGCAAG	1348
Db	128	ATATGACATCAGCGCTGTATTAAGAAAGGACAGAGATGTTGAATACGTTGTTTGAG	187
OY	1349	AAATATGCGAGAAATGTACATGTGTTCAAAACAAAACCTGGCTGAAAGTTTCAACAG	1408
Db	188	AGAGATGGAAGG-----AAACCTGTCTAATTAATAATTTTAA	227
OY	1409	AATTAATTAAGTATGTGAGAACTGAAAAAGAAAAAAGAAAGAAATTTCTCAACGATG	1468
Db	228	TGATATCTGAAGATTTGCAAAAATATCAGTGGCAAAAATGTCAGAACTGTATTAAGATA	287
OY	1469	AATTAATTAATTCGTGTGTACAGCAGCAAAAGACGCCGGTGTCTACACATGATCTTC	1528
Db	288	CTTTCAAAATGTCTTT---ACCTTAAGACTATGTATGTCTTTCAGATATATTT	343
OY	1529	GAATGAAAACATCTTTTATACGACAACTGATCAAAAACGAGATTTCCCGACAGATA	1588
Db	344	TTCTTCATCAAAAGAGTTAAGTGTGCTTTAGATGATCAGAGAGATTTTTCATTAGAAA	403
OY	1589	AAATTTGCAGGAATTTGGGAGAAAGTCCAGATTTTAGAGAGAGATTTCAAAAGAAATTA	1648
Db	404	AGGATTTGTTGAAATTTGGAGAGAAAGTGTATGAATCTGTAGTATCTATATTTGAAAT	463
OY	1649	CATGCGCATGTCATACACTGGAGAGCAATGCATCGCTTGGGACATACAGAAATTTTAA	1708
Db	464	TAGAAAGTGTATTAACATTAAGAAAGCGCTGTGAATCTTTAAGTTACGAAAGATTTA	523
OY	1709	AGCAGGTTTATTTGATGAACACAAAGATCTTTGAAAGACCAAGAAAGTTGTCTAAAT	1768
Db	524	GAAAGATATTTTATGAAAGAAAGAAATATCACTTATTCGATGAGCAAAATTTGACGAGG	583
OY	1769	ACCTTAAGAAAGAAAGCTGTATTAATTAATGCTGTGAAGAGAGATGACCGTTTCTTTGTAT	1828
Db	584	CGTTGCATGAAATGTGAAGCTTTATTTAGAAAGAGAGAAATCCATTTGAGTTTTCAT	643
OY	1829	GTCGCTTCCAAAACGCTACGTGTAGCTGTATGTAAAGACGTGAAGACAGGTGTGAAG	1888
Db	644	GTGCTTTGCAAGGAAGAAACATGTCAACGTATGTATCCATACAACTCAAGATTTGTATTT	703
OY	1889	TATTTCAAAAAAATATTAAGCTTCATATATATATTAATTTCTTG-----AAATA	1939
Db	704	ATTTTAAGACACACATCAAAAATAAAAAATTTAGAAACAAATTTGAAAGTAACAGG	763
OY	1940	ATACAAATTAATAATTAACAACACTGGAAGAAATTTCCCTCTTGCGATACGATTTGCATA	1999
Db	764	ATTAATCAAAAGAGACGAAGTATAGAACCTGTGCACAAACATGGGGTAAATATTTGTCAAC	823
OY	2000	GATTTTCACTAATTTGTCTCAGGCTTTACGAAGAGAAATGTTGTACAAAAATCAAGAAC	2059
Db	824	AACTTATGSAATATGTTCCAGATATGTTGAAAAAAGAAAAAAGAACATGTACATA	883
OY	2060	ATCGTGAAGCGTTCTATTAAGAAAGAGCGTTGGAAGATGCTCAAGTGAAGCTTCAAG	2119
Db	884	ATCAAAACTGGAAGAACTCGAAAAAATATGCACTGATACCTTTAAA-----AGTTGGA	939
OY	2120	GAAATTTGACTGATTAATCTTAATATGTGAACCTGCAATTTGAAGAAAGATTTGTACAGTACGG	2179
Db	940	TTGAAGATGAGCTGACATCTGTTGAAGGACAGCTTTAAAGATTAAGAAATATGTAAA	999
OY	2180	GAAACGTAAATATATGCGTCATACAGGTGCTTATGCAAAAGCTAAACAGAGATTAACCTG	2239
Db	1000	GTAACCTAGGACAGCGTGTCCCTGAAGTTGAATAATATGATATTCATAAATTTCTGT-	1058

OY 2240 GAAAGTGTATGAGATGCTAGAAAGAACTCTGTGAGAAATTAGTGAAGAAGTGGAG 2239
Db 1059 -TACTTAATTGTGAAGATTCCTTGGAAAAATGTTGCGCGAAATTAGTTAAAAAAGTACAGA 1117
OY 2300 AACAGTGCAGAACATTACCAACAGAAATTAGACCAACCGCAGCTGATCTAAAAAAGATT 2359
Db 1118 AGAAATGTCTTACTTTTAAAAAGCACTGATTAAGCCGAAGATGATGACCAAGATGA 1177
OY 2360 ATTAAGACATATGAGGAAGCTTAAAGAAAGTGCAGAGAAAGCAATGACAAAGTCCAGTCTTG 2419
Db 1178 AGACTGAGTAGCAAAAAATGCTAAAAAGCGGCGAGAAAGATCTACAAACAAAGCTAGCTTAT 1237
OY 2420 TTTTGTCACTCTTAAAGAAAAAGCAA 2446
Db 1238 TGTATCAAAAGTCTGAAAAAGCCGCA 1264

Search completed: October 29, 2005, 02:45:56
Job time : 1582 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 29, 2005, 01:23:45 ; Search time 9452 Seconds
(without alignments)
12371.290 Million cell updates/sec

Title: US-10-654-416-13
Perfect score: 3072
Sequence: 1 atgagcgagcggtcgaacgc.....ttattcgttcacgatttag 3072

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68473088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hnc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_g881:*
9: gb_g882:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	132.6	4.3	737	2	AM334880
C 2	128.6	4.2	1811	2	CG753732
C 3	121.4	4.0	719	2	AM334763
C 4	119.8	3.9	1531	9	CG748014
C 5	116.8	3.8	2001	9	CG748014
C 6	116.4	3.8	1784	9	CG748014
C 7	115.2	3.8	1981	9	CG748014
C 8	115	3.7	1727	9	CG748014
C 9	114.6	3.7	1453	9	CG748014
C 10	114.4	3.7	1256	9	CG748014
C 11	111	3.6	1491	9	CG748014
C 12	110.2	3.6	1843	9	CG748014
C 13	109.4	3.6	699	2	AM333354
C 14	108.8	3.5	1632	9	CG748014
C 15	108.6	3.5	1377	9	CG748014
C 16	108.4	3.5	1434	9	CG748014
C 17	108.4	3.5	1797	9	CG748014
C 18	108.4	3.5	2061	9	CG748014
C 19	108.2	3.5	1519	9	CG748014
C 20	108.2	3.5	1519	9	CG748014
C 21	107.8	3.5	1477	9	CG748014
C 22	107.6	3.5	1407	9	CG748014
C 23	107.2	3.5	1608	9	CG748014
C 24	106.8	3.5	1736	9	CG748014

C 25	105.8	3.4	1376	9	CG747831
C 26	105.6	3.4	1829	9	CG747831
C 27	105.6	3.4	1863	9	CG747831
C 28	105.4	3.4	1641	9	CG747831
C 29	105.2	3.4	1716	8	CG747831
C 30	105	3.4	1594	9	CG747831
C 31	105	3.4	1824	9	CG747831
C 32	104.6	3.4	2030	9	CG747831
C 33	104.6	3.4	1599	9	CG747831
C 34	104.4	3.4	1842	9	CG747831
C 35	104	3.4	1416	9	CG747831
C 36	103.8	3.4	1533	9	CG747831
C 37	103.8	3.4	1536	9	CG747831
C 38	103.8	3.4	1808	9	CG747831
C 39	103.8	3.4	1823	9	CG747831
C 40	103.8	3.4	1506	9	CG747831
C 41	103.4	3.4	1506	9	CG747831
C 42	103.4	3.4	1276	9	CG747831
C 43	103.2	3.4	1654	9	CG747831
C 44	103.2	3.4	1757	9	CG747831
C 45	103.2	3.4	1757	9	CG747831

ALIGNMENTS

RESULT 1
LOCUS AM334880/c 737 bp mRNA linear EST 31-JAN-2000
DEFINITION S40C12 AGS-1 Pneumocystis carinii CDNA 3', mRNA sequence.
ACCESSION AM334880
VERSION AM334880.1 GI:6831237
KEYWORDS EST.
SOURCE Pneumocystis carinii
ORGANISM Pneumocystis carinii
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 737)
Smilian, A.G., Arnold, J., Weise, M., Wunderlich, J., Staben, C.,
Edman, J.C., Kovacs, J. and Cushman, M.
Expressed sequence tags from Pneumocystis carinii
Unpublished (2000)
CONTACT: Staben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: staben@pop.uky.edu.
AUTHORS location/Qualifiers
1. 737

FEATURES

source
1..737
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_id="AGS-1"
/note="Vector: Lambda ZAP II; Site 1: EcoRI, Site 2: XhoI;
P. carinii organisms (3x10e9) from a single rat (99-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Tritol extracted RNA. Oligo dt priming, standard
conditions described by vendor. StrataGene. Further
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN

Query Match 4.3%; Score 132.6; DB 2; Length 737;
Best Local Similarity 52.3%; Pred. No. 1.7e-16;
Matches 396; Conservative 0; Mismatches 319; Indels 42; Gaps 3;

1657 TGTATACACTGAGCGAGCAATCGCTGGGAGCTACAGAAATTTAAAGCAGGTT 1716
737 TGTATTAATTTGAATGATGACGCTGCAAAATTTAGAAAGTCTCAGAAAGATCAAGACATA 678

	QY	1717	TTATGGATGTAACCAAGATACTTGTGAACAACAAGAAAGTGATTAATACCTAAA	1776
	Db	677	TTTTTAAAGAAGAAAAAGATTATTAAGAAATCAAGAAATGTGAAAAGGCTCTGAA	618
	QY	1777	GAAAAAGTAAATAAAGTCTTAGAAGAGAGATGACCCTTTCTCTTGTATGTGCTTC	1836
	Db	617	GAAAAAATGTATGGAAATTTTAAAAAGAGCAAAATCTATTAGATTCTCATGTGTTTA	558
	QY	1837	CAAAACCGCTACGTGTAGCTGATGTGTAAAACACGTGAAAAGCAGGTGTGAAGTATCAA	1896
	Db	557	CAGAAAAGAAATATGTCCGACTATATGTCMAAGACAGAAATTCAGTGCATATTTCTTAA	498
	QY	1897	AAAAATATPAAAAGCTCATATATTATTAATTTCTTGGAATAATATPACAAATPAAATACA	1956
	Db	497	GAGAACATAGAAAAACAGAAATTTTGA---TAAAGAAAACGAAAAAGAAATGAAACA	441
	QY	1957	ACACTGGAAAAGAAATGTCCCTCTTGCGATACGATTTGCAATAGATTTTCACTAATGT	2016
	Db	440	TCACTTGAAAGAAATATGTTTTATTTATGAGTGCATATTGCATTCAGCTTATGAAAAATTGT	381
	QY	2017	CCAGGTCTTACGAAAAGCATATG-----TTGTACAAAATCAAGAG	2058
	Db	380	CCAAAGCTATTGAAAAGAAAGATPATACTGATGAGATTGTGTGAAACTTPAAGAG	321
	QY	2059	CATCGTAGCGCTTCTPAAAAGAAAGCCCTGGAAAGTGTCTCAAAAGTAGAGCTTCAA	2118
	Db	320	AATATGACCTATTTCTGGAAAAGAGACCTTGGATGATGATTAATGCTCATAAATTGAA	261
	QY	2119	GGAATATGACTGATPAAATCTPAAATGTGAACTGCATTGMAAAGATATTGTACAGTACG	2178
	Db	260	GGAATTTAATAGTGAATATGCTPAAATGTGTAATAGGTTAGAGAGCTATTCACATCAATG	201
	QY	2179	GGAACCTAAATATGCGTCAATCAGTGGCTTATGCAAAAGCTAACCAAGATPACTCT	2238
	Db	200	AAAACGTGAACGAATPACTTTAAAAAATCAGTATATCTGATGAAAATAAGAGC---	144
	QY	2239	GGAAGAAGTATGATGAGATGCTAGAAAGGAACTGTGTGAATTAAGTAAAGAAAGTGSAA	2298
	Db	143	-----GTTCAAGATGAATTTGTGAAAGATTATGTAACAAAAATAGCG	102
	QY	2299	GACAGTGCAAAGCATTAACAACAGAAATTAGACAACCGCACGTGATCTAAAAAAGAT	2358
	Db	101	AAAGAAATGCCAATCTPAAAAATTAATTTABAAAAAGCGGAAAAAAGATTTCAGACAAA	42
	QY	2359	TATPAGACATATGAGAACTTAAGAAAACGTGCAGAG	2395
	Db	41	GAAAAAGATATTAAAAAAAAAAAAAAAAAAACTCGAG	5
RESULT 2				
CG753732/c				
LOCUS			1811 bp	DNA linear GSS 24-OCT-2003
DEFINITION			P048-f-G03.ya Ppa EcorI BAC library Pristionchus pacificus genomic,	
VERSION			CG753732	
KEYWORDS			CG753732.1 GI:37978509	
SOURCE ORGANISM			GSS.	
			Pristionchus pacificus	
			Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;	
			Neodiplogasteridae; Pristionchus.	
REFERENCE			1 (bases 1 to 1811)	
AUTHORS			Strinivasan,U., Sinz,W., Jesse,T., Wiggers-Berebolle,L., Jansen,K.,	
			Buntjer,D., van der Meulen,M. and Sommer,R.J.	
			An integrated physical and genetic map of the nematode Pristionchus	
			pacificus	
TITLE			Mol. Genet. Genomics 269 (5), 715-722 (2003)	
JOURNAL				
MEDLINE				
PUBMED				
COMMENT				
			Contact: Sommer RJ	
			Max-Planck-Institute for Developmental Biology	
			Spemannstr. 37-39, Tuebingen D-72076, Germany	

FEATURES
 source location/Qualifiers
 1. 1811 "rd44f-cabbus 2

```

/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strains="california"
/db_xref="taxon:54126"
/clone_1lb="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

```

Query Match	4.2%;	Score 128.6;	DB 9;	Length 1811;
Best Local Similarity	35.9%;	Pred. No. 1.2e-15;		
Matches 509;	Conservative 0;	Mismatches 899;	Indels 0;	Gaps 0;
QY	72	AAATTCTAAGAGAGATGCACTAAGTGCACGAAATGCACAAAAAACTTAATAAATATTG	131	
DB	1585	AAA	1528	
QY	132	CCAAGATTCAGTACGACCAAACTTAATATTGACCAAGTACACAGAAAACCTTAAGCTTT	191	
DB	1525	AAA	1466	
QY	192	TTGCGAAGATGGAAGACAGATACAAAATGCAAGAACTGAAGCCAAATTGAGAAAAA	251	
DB	1465	AAA	1400	
QY	252	ATGTATCTACAACTCAAGAGAACTTAAGAGCAATTAABAAAAAAATTCAGATTATAC	311	
DB	1405	AAA	1344	
QY	312	GGATTAAGATTGCAAGAGAGATGACAACATGCTCTTTTGGAGGAGTATGTTCAA	371	
DB	1345	AAA	1288	
QY	372	AGAACTTAAGATGATTCCAATCTTGAAGATTAAGTCTATCAAAAGAAACGTGATPA	431	
DB	1285	AAA	1228	
QY	432	AGTTGCGGAGAAAGTTCTTTAAGAGACTTCGTAGGAGTCTTAATGATCAGTCAATG	491	
DB	1225	AAA	1166	
QY	492	TGAAAAAAAACTTAAGAGATTTGCCCTGTCATGGAGGAGGAAGTGAATTACAA	551	
DB	1165	AAA	1100	
QY	552	CTTGTCGTGAACGAGAAAGACATGTAAAGATATTTTAAATTGAAAAAGTAAAGTGTG	611	
DB	1105	AAA	1044	
QY	612	CGGTACTCTTAAAAAGATGTTTCAGCAGCACTGGAAGTTTAAAAAAGAAACATGTCT	671	
DB	1045	AAA	986	
QY	672	TGAATTACTGAACTAATCTATTTTTCATTTGAATTCGAGACGACGATATATTAA	731	
DB	985	AAA	926	
QY	732	ATGTATTTGAATTTGGAGAGAAAATGCCAAGACAAATATTGCTTATATGCCACGAGAC	791	
DB	925	AAA	866	
QY	792	CGATTTTGATCCAACTAGCCAGAGCTACATATGACAGAGATATTAGGCTGGAAGAGTT	851	
DB	865	AAA	806	
QY	852	TTAATAGAGGTAGAGAGATGAGATTTTATTGAAAGATCATCTTAAGATGCGAC	911	

Max-Planck-Institute for Developmental Biology
 Spemannstr. 37-39, Tuebingen D-72076, Germany
 Tel: 00497071601371
 Fax: 00497071601498
 Email: ralf.bommer@uebingen.mpg.de
 Class: BAC ends.

FEATURES

Source
 1..1531
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strain="California"
 /db_xref="taxon:54126"
 /clone_lib="ppa EcORI BAC library"
 /note="The library was generated by a partial digest of
 the genomic DNA with EcoRI and cloning into the BAC
 vector."

ORIGIN

Query Match 3.9%; Score 119.8; DB 9; Length 1531;
 Best Local Similarity 38.6%; Pred.No. 7.7e-14;
 Matches 523; Conservative 0; Mismatches 832; Indels 0; Gaps 0;

```

Oy 93 AAGTAGCAGGAAATGCAAAAATTAATATGTCAGAGAAATGCTGAGCAAA 152
Db 177 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN 236
Oy 153 ACTAATATAGAACAGTACACAGAAAATTAAAGTTTTCGAGATGGAAGCAGA 212
Db 227 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 296
Oy 213 TACAAAATGCAAGAACTGAAAGCCAAATTGAGAAAAATGCTACATCAAGGAA 272
Db 297 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 356
Oy 273 ACTTAAGAGCAATTAATAAAAAAAAAATTCAGATTTAAACGATAAGATTGCAAGAGA 332
Db 357 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 416
Oy 333 TGAACACAAATGCCATATTTTGGAGGAGTATGTTCAAAAGAACTTAAGATGATGCA 392
Db 417 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 476
Oy 393 TACTTGAGAAATAGTGCTATCAAAAGAACTGATAAAGTTGGGAGAAAGTTCTTT 452
Db 477 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 536
Oy 453 AAGACACTTCGTAGCGATCTTAATGAGTCACTCATGTGAAAAAACTTAAGAGAT 512
Db 537 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 596
Oy 513 TTGCCCTGTCAATGGGAGGAGAAAGTATGATTAACAACTTGCTGGAACAGAAAGA 572
Db 597 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 656
Oy 573 GACATGTAAGATATTTTAATGAAAAAGTAAGAGTCGGTACTCTTAACACAGATG 632
Db 657 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 716
Oy 633 TTCACAGACACTAGGAAGTTTAAAAAAGAAACATGCTTTGAATTAAGTGAATGGAGAAA 692
Db 717 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 776
Oy 693 TTTTACATTTGGAATTTGGAGAGACGATTAATTAATGATTTGAATTTGGAGAAA 752
Db 777 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 836
Oy 753 ATGCCAAGAACTAATATTGCTTAATATGCAACAGAGCCGATTTTGATCCAATGAGCC 812
Db 837 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 896
Oy 813 AGAGGCTACATATAGCAGAGATATAGGCTGGAAGGTTTAAAGAGGTGAGAGAGA 872
Db 897 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 956

```

```

Oy 873 TGAATTTTAATGGAAGAAATCATCTAAGAGATGCGACAGCTTTGTTGGCATTTGAT 932
Db 957 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1016
Oy 933 CCAAGATTTAGTGTCTTAATAAAGAGCAGCAAGAGAAATGCGAGAGCCCTTCAAA 992
Db 1017 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1076
Oy 993 AAGCTGCAAAATTCCTCATGACATGAGCGCTTTAGAAAGTTATGTAGAGAAAAATGCTTT 1052
Db 1077 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1136
Oy 1053 AAGTAATGATGGAAGCAAAAAATGTGAAGATTGCAAAATGATATTAACAATACTTGCA 1112
Db 1137 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1196
Oy 1113 AATTTGACTTCAAAGTCACTAATAATCGTCTTTTGATCCCAAGAAATTAATGA 1172
Db 1197 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1256
Oy 1173 AATTGTTGATGGAGAGGTTGCCACATTTCTTAGCAACGAGATTGTGCAAAATTGGA 1232
Db 1257 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1316
Oy 1233 GTCTTATTTGTTCTAATTTTGAATAAATGTCAGATGAGAGAAATGATTAATAATAT 1292
Db 1317 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1376
Oy 1293 AAGACCAATGTTACAAAGAGAGACTTGATGACGCGCAATTAAGTCTGCAAGAAA 1352
Db 1377 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1436
Oy 1353 TATCGAGATGTTTACATGTTCAACAAAGCTGCTGAAAGTTTCAACAGAAAT 1412
Db 1437 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1496
Oy 1413 AGTAAAGTATGTGAGAACTGAAAAAGAAACA 1447
Db 1497 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1531

```

RESULT 5
 Cui18787
 LOCUS
 DEFINITION
 ISB1-72M23.T7.2 ISB1 Xenopus tropicalis genomic clone ISB1-72M23,
 genomic survey sequence.
 Cui18787
 ACCESSION
 Cui18787.1 GI:40612422
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 GSS.
 Xenopus tropicalis (western clawed frog)
 Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 2001)
 Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
 Mardis,E. and Wilson,R.
 TITLE
 JOURNAL
 COMMENT
 Unpublished (2003)
 Contact: Richard K Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@wustl.edu
 Insert Length: 75000 Std Error: 0.00
 Seg primer: T7 TAATACGACTCACTATAGG
 Class: BAC ends
 High quality sequence start: 384
 High quality sequence stop: 783.
 Location/Qualifiers
 1..2001
 /organism="Xenopus tropicalis"
 /mol_type="genomic DNA"

/db xref="taxon:8364"
/clone="ISB1-72M23"
/note="Vector: pBelobAC1; ISB-1 Xenopus tropicalis BAC
Library Segment 1"

ORIGIN

Query Match 3.8%; Score 116.8; DB 9; Length 2001;
Best Local Similarity .44.8%; Pred. No. 3.2e-13;
Matches 620; Conservative 0; Mismatches 752; Indels 12; Gaps 4;

QY AAAAGAGATGAGCTAAGTGCAGAGATGCAAAAAAATTAATAATATGTCACAGA 137
DB AAG 607
QY 138 ATTGATGAGCAAACTAATATATGACAACTACAGAAAACTTAAGCTTTGCGA 197
DB AAGA 667
QY 198 AGATGAGAAAGCATACAAATGCAAGAACTGAAGCCAAATTGAGAAAAATGTAC 257
DB AAGA 727
QY 258 TACATCAAGAGAACTTAAGAGCAATTAATAAATAATTCAGATTACGATAA 317
DB AAGA 787
QY 318 GATTCGAAGAGATGACACAACTCTATTTTGGAGGAGTATGTTCAAAAGACT 377
DB GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAATTAATTAAGAAAGAAAAAATAA 847
QY 378 TAAAGATGATTCGAATCTTTGAGAAATTAAGTCTATCAAAAGAAAGCTGATTAAGTGC 437
DB AATTAAGCAAAATTAATAAATAAATAAAGAAAAAATAAAGAAAAAAG 907
QY 438 GGAAGAGTTCTTTAAGAGCACTTCGAGCATCTTAATGATCACTATGATGAAA 497
DB AATTAATTAATAAATAAATAAATAAAGAAAAAATAAAGAAAAAAG 967
QY 498 AAAAATTAAAGAGATTTCCCTGTCAAGGAGGAGAAAGATGATTAACAACTGTG 557
DB AAAAAATTAAGAAATTAATAAATAAATAAAGAAAAAATAAATTAATTAATTAATTA 1027
QY 558 TCTGAACGAGAAAGACATGTAAGTAATTTTAATTAAGAAAAAGATGAAGTGGTAC 617
DB AAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1087
QY 618 TCTTAAA--ACAGATGTTTCAGACACTAGAGAGTTTAAAAAAGAAACATGCTTGA 675
DB AAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1147
QY 676 TTAATCGAAGATCTATTTTACATGGAATTCGCGAGACGATATTAATTAATGT 735
DB AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1204
QY 736 ATTGAATTTGGAGAAATGCGCAAGAAATAATGCTTATATGCCACGAGCCGAT 795
DB AGATAATACAAAAAAGAAAAAAGAAAAAAGAAATAGAAAAAAGAAAAAAGAAAA 1264
QY 796 TTGATTCAACTAGGCGAGAGCTACATAGCAGAGATATAGGCTGAGAGATTTAT 855
DB AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1324
QY 856 AAGAAAGTGAAGAGATGAGTTTATTTGAGAAAGATCATCTTAAGATGCAAGCT 915
DB AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1384
QY 916 TTGTTGCGATTGATCGAAGATTTCTAGCTTAATAA-----AAAGACGACAAAG 969
DB AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1444
QY 970 AATGCGAAGAGCCCTTCAAAAAAGCTCAAAATCTCTATGAACATGAGGCTTTAGAA 1029

DB 1445 AAAAAAAAAAGAAATATAGAAAAAAAAAAAAAAAAAATTAAGAAAAATTAAGAAAAA 1504
QY 1030 AGTTATGTAGAGAAATATGCTTTAAGTAATGATGAAACGAAAAAATGTAAGATTGCA 1089
DB 1505 GAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1564
QY 1090 AATGATATTAAACAAAATTGCAAAATTTTCACTTCAAAAGTCACTAATAATCGTCTTTT 1149
DB 1565 AAAAAAAAAAAAAAAAAAGAAAAATTAATAAATAAATAAATAAATAAATAAATAAATA 1624
QY 1150 GATCCAAACAAAGAAATTAAGAAATTTGTTGATGAGGAGGTTGCCACATTTCTTAC 1209
DB 1625 AAT-GAAGAGAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1683
QY 1210 AACGAAGATTGCGAAATTTGAGTCTATTTGTTCTATTTGAAAAAATAATGTCAGAT 1269
DB 1684 CAAAAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1743
QY 1270 GGAAGAAATGCGATGTAATAATTAAGACCAATGTTACAAAAGAGACTTGATGACGG 1329
DB 1744 AAGAAAAAGAAAAGAAAAAATAAAGAAAGAAATTAATAAATAAATAAATAAATAA 1803
QY 1330 GCAATTAAGTCTGCGAAGAAATTAAGCAGAGATGTTACATGTTCAACAAAGCTG 1389
DB 1804 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1863
QY 1390 CTGAAAGATTTCACAGAAATTAAGTAAGTATGTAAGAACTGAAGAAAAAGAAACAA 1449
DB 1864 AAGAAAAAGTAAAGAAATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1923
QY 1450 GGA 1453
DB 1924 CAAA 1927

RESULT 6

CL081992

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

MARDIS E. and Wilson R.

A physical map of the xenopus tropicalis genome

Unpublished (2003)

Contact: Richard K Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submisions@wustl.edu

Insert Length: 175000 Std Error: 0.00

Seq primer: RM4 ctcaaggcatcgctgagc

Class: BAC ends

High quality sequence start: 409

High quality sequence stop: 503.

Location/Qualifiers

1..1784

/organism="Xenopus tropicalis"

/mol_type="genomic DNA"

/strain="Nigerian frog"

/db_xref="taxon:8364"

/clone="CH216-165M9"

/sex="male"

/cell_line="Stock 248 F7A2, inbred N7"

/note="Vector: pTARAC2.1; CHORI-216 Xenopus tropicalis"

1784 bp DNA linear GSS 31-DEC-2003
CH216-165M9, RM4.1 CH216 Xenopus tropicalis genomic clone.
CH216-165M9, genomic survey sequence.
CL081992.1 GI:40537905
GSS.
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 1784)
Kremetzki, C., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.


```

Db 1700 ACTAAAAAGAAAAATATATAAATCTTAAAAAAGAAAAAAG 1641
Qy 164 AACAGATACAGAAAACTTAAGGTTTTCGAGATGAGAAAAGCATACAAAAGCA 223
Db 1640 AAAATATATATATATAAATAATATATATATATATATATATATATATATATAT 1581
Qy 224 AAGATGAAAGCAATATTGAGAAAAATGTAATCAATCAAGAAAGAACTTAAAGAG 283
Db 1580 AAAATATATATATATATATATATATATATATATATATATATATATATATAT 1521
Qy 284 CAATTAATAAAAAATTCAGATTATACGATTAAGATTGCAAGAGATGAAACAAAT 343
Db 1520 TAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 1461
Qy 344 GCGATTTTGGAGGAGTATGTTCAAAAGAACTTAAAGATGATGCA-----ATCTT 397
Db 1460 ATATATATATATATATATATATATATATATATATATATATATATATATAT 1401
Qy 398 TGAGAAATTAAGTCTATCAAAAGAAAGTATTAAGTTGCGAGAAAGTCTTTTAAGAG 457
Db 1400 TAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 1341
Qy 458 CACTTCGATCGATCTTAATGATCAGTCAATGTGAAAAAATTTAAAGAGATTGCC 517
Db 1340 TAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 1281
Qy 518 CTGTCATGGGAGGAAAGTATGATGATTAACAACTGTGTCTGAACCAAGAGATCAT 577
Db 1280 AAAATATATATATATATATATATATATATATATATATATATATATATATAT 1221
Qy 578 GTAGAAATTTTATTAAGAAAGATGAGAGCGGATCTTAAACAGATGTTTCAG 637
Db 1220 AAAATATATATATATATATATATATATATATATATATATATATATATATAT 1161
Qy 638 CAGCACTGAGAGTTTAAAAAAGAAACATGTCTTGAATTAATCAATGCTATTTT 697
Db 1160 TAAATATATATATATATATATATATATATATATATATATATATATATATAT 1101
Qy 698 ACATTGGAATTCGGAGACAGCATTAATTAATGATTAATGGAGGAAAGATGCC 757
Db 1100 AAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 1041
Qy 758 AAGACAAATATATGCTTATATGACACAGGACCGGATTTTGAATCAATGAGCCAGAG 817
Db 1040 AAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 983
Qy 818 CTACATATGACAGATATAGGCTGAGAGTTTATAGAGGTAGAGAGATGAG 877
Db 982 ----TATATATATATATATATATATATATATATATATATATATATATATAT 927
Qy 878 TTTTATATGAGAGATCATTAAGATGCGACAGCTTTGTCGATTTGATCCAG 937
Db 926 AAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 867
Qy 938 ATTCTAGTCTTAAAAAAGAGACAAAGAGAAATGCAAGAGCCCTCAAAAAAGCT 997
Db 866 AATATATATATATATATATATATATATATATATATATATATATATATATAT 807
Qy 998 GCAAAAAATCTCATGACATGAGGCTTGAAGATTTATGTAAGAAAAATGCTTAAGTA 1057
Db 806 ACAAAATATATATATATATATATATATATATATATATATATATATATATAT 747
Qy 1058 ATGATGAGACGAAAAATGAGAGATTGCAAAATGATATTAACAACCTTGCAAAATTT 1117
Db 746 TAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 687
Qy 1118 TCACTTGAAGAGTAAATATATATATATATATATATATATATATATATATATAT 1177
Db 686 AATATATATATATATATATATATATATATATATATATATATATATATATAT 627
Qy 1178 TTGATGAGAGAGGTTGCCAATTTCTTATGCAACGAAGATTTGCAAAATGAGTCT 1237

```

```

Db 626 AAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 567
Qy 1238 ATTGTTTCTATTTTGAAGAAAAATGTCAGATGAGAAATGATGTAATATATAGAG 1297
Db 566 AAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 507
Qy 1298 CAACATGTTCAAAAGAGACCTGATGACGCGCAATTAAGTCTGCAAGAAATATGTC 1357
Db 506 AAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 447
Qy 1358 GAGCATGTTTACATGTTTCAACAAAGCTGCTGAAAAATTTCAACAGAAATTAATA 1417
Db 446 AAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 387
Qy 1418 AAGTATGTCGAACTGAAAGAAAGAAACAAAGAA 1453
Db 386 AAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 351

RESULT 8
LOCUS CL114085 1727 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-59011.Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-59011,
ACCESSION CL114085
VERSION CL114085.1 GI:40607720
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1727)
Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
A physical map of the xenopus tropicalis genome
Unpublished (2003)
CONTACT: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submis@wustl.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTAGTGACACTATAG
Class: BAC ends
High quality sequence start: 280
High quality sequence stop: 1423.
Location/Qualifiers
1. 1727
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-59011"
/clone_1b="ISB1"
/notes="Vector: pBeLoBAC11; ISB-1 Xenopus tropicalis BAC
Library Segment 1"

ORIGIN
Query Match 3.7%; Score 115; DB 9; Length 1727;
Best Local Similarity 42.3%; Pred. No. 7,4e-13;
Matches 601; Conservative 0; Mismatches 819; Indels 0; Gaps 0;

Qy 34 ACACAGATGATGATGAGAAACATGTTTATGCTTATATCTAAAGGAAGATGCACTA 93
Db 165 AAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 224
Qy 94 AGTACGACGAGATGCAAAAAATTAATAATATATGTCAGAAATGACTGAAGCAAAA 153
Db 225 AAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 284
Qy 154 CTAAATTTAGACAGATGACAGAAACCTTAAGGTTTTCGAAAGTGAAGCAAGAT 213
Db 285 AAAAAAATATATATATATATATATATATATATATATATATATATATATATAT 344

```

```

Oy 214 ACAAATGCAAGAACTGAAAGCCAAATTGAGAAAAATGTACTCAATCAAGAGAAA 273
Db 345 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 404
Oy 274 CTTAAGAAGCAATTAAAAAATTCAGATTATACGATTAAGGATTGCGAAGAAAT 333
Db 405 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 464
Oy 334 GAACAAATGCTATTTTGGAGGAGTATGTTCAAGAACTTAAAGATGATTGCAT 393
Db 465 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 524
Oy 394 ACTTGAAGAAATTAAGTCTATCAAAAGAACTGTAAAGTTGGCGAAGAACTCTTTA 453
Db 525 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 584
Oy 454 AGAGCACTTCGATCGATCTTATATGATCATATGATGAAAGAAAACTTAAAGAAAT 513
Db 585 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 644
Oy 514 TGCCCTGTCATGGGAGGAGAAAGTATGATTACAAACTGTGTCTGACCCAGAAAG 573
Db 645 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 704
Oy 574 ACATGTAAGATATTTTATTAAGAAAGATTAAGAGTGCCTCTTAAACAGATGT 633
Db 705 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 764
Oy 634 TCAGCAGCACTAGAGAGTTTAAAAAAGAAACATGCTTGAATTTCTCGAACATGCTAT 693
Db 765 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 824
Oy 694 TTTTACATTGGAATTGCGAGACGATATTAATTAATGATTGAATTGGAGAGAAA 753
Db 825 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 884
Oy 754 TGCCAGAAACAAATATGCTTATATGCGACACAGACCCGATTTGATCCAACTAGGCA 813
Db 885 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 944
Oy 814 GAGGCTCAATAGCAGAGATATAGGCTGGAAGATTATTAAGAGATGAGAGAGAT 873
Db 945 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1004
Oy 874 GGAGTTTATTTGAAAGATCATCTAAAGATGCGACAGCTTTGTGGCATTTGTGATC 933
Db 1005 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1064
Oy 934 CAAGTTCTAGCTTAAATTAAGAAAGCAAGAAAGAAATGCGAAGAACCTTCAAAA 993
Db 1065 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1124
Oy 994 AGCTCAAAAAATCCTCATGAACATAGGCTTTAGAAAGTTTATGTAAGAAAAATGTTTA 1053
Db 1125 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1184
Oy 1054 AGTAATGATGAAACGAAATATGTGAAGATTGCAAAATGATATTACAAAATTGCAA 1113
Db 1185 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1244
Oy 1114 ATTTTCATCTCAAAAGTCACTAATATGCTCTTTTGTCCAAACAAAGAAATTAAGA 1173
Db 1245 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1304
Oy 1174 ATTTGATGAGGAGGCTGCAACATTTCTTACCAAGAAAGTTGCGAAATTTGAG 1233
Db 1305 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1364
Oy 1234 TCTATGTTTCTATTTGAAAAAAATGTCAGATGAGAAATGCAATGTAATAATATA 1293
Db 1365 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1424
Oy 1294 AGAGCAATGTTCAAAAAGAGACTTGATGACGCGCAATTAAGTCTGCAAGAAAT 1353

```

```

Db 1425 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1484
Oy 1354 ATGCGAGAAATGTTACATGCTTCAACAAAGCTGCTGAAAGTTTCAACAAATTA 1413
Db 1485 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1544
Oy 1414 GTAAAGTATGTGAAACCTGAAAAAGAAACAAAGAA 1453
Db 1545 ACAAACACAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1584

RESULT 9
AJ591978/c 1453 bp DNA linear GSS 15-JAN-2004
Arabisopsis thaliana T-DNA flanking sequence, right border, clone
598D01, genomic survey sequence.
AJ591978
AJ591978.1 GI:37941602
GSS: right border; T-DNA flanking sequence.
Arabisopsis thaliana (thale cress)
Arabisopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
1 Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Samson, F.,
Chauvin, S., Bechtold, N., Cruaud, C., Dehose, R., Pelletier, G.,
Lepoint, L., Caboche, M., and Lecharny, A.
T-DNA integration into the Arabidopsis genome depends on sequences
of pre-insertion sites
EMBO Rep. 3 (12), 1152-1157 (2002)

JOURNAL
MEDLINE
22363535
PUBMED
12446565
REFERENCE
2 (bases 1 to 1453)
Balzergue, S.
Direct Submission
Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue
Gaëtan Cremieux, 91057 Evry cedex, FRANCE
PCR was performed on DNA from transformants of Arabidopsis thaliana
plants from INRA (Versailles). The DNA fragment(s) resulting from
the PCR were directly sequenced from the left or the right border
to determine the genomic sequence flanking the insertion. T-DNA
derived sequences were removed. Information to order the
corresponding mutant line and a link to a database providing a
graphical display of the insertion site are available at
http://dbsegap.versailles.inra.fr/publications/. This sequence has
been generated in the framework of the French plant genomics
program 'Genoplante' (http://www.genoplante.com and
http://genoplante-info.inbio.gen.fr).

FEATURES
source
1..1453
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Wassilewskija"
/db_xref="taxon:3702"
/clone="598D01"
/clone_id="Arabidopsis thaliana T-DNA insertion lines"
1..1453
/misc_feature
/notes="T-DNA flanking sequence
right border"

ORIGIN
Query Match 3.7%; Score 114.6; DB 9; Length 1453;
Best Local Similarity 38.9%; Pred. No. 8.9e-13;
Matches 544; Conservative 0; Mismatches 852; Indels 3; Gaps 1;

Oy 79 AAGAAAGTGAAGCTTAAAGTGAAGCAAGAAATGCAAAATTAATGCGAAGAA 138
Db 1427 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1368
Oy 139 TTGACTGAAGCAAACTTAATATATGAACAGTACAGAAACTTAAAGTTTGGCAA 198

```

[illegible][illegible]

```

QY 258 TACAATCAAGGAATCTTAAAGAACTTAAATAAATTCAGATTATACGATTA 317
DB 1377 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1318
QY 318 GGATTGCAAGGAATGAACAATGCGCTATTTTGGAGGAGTATGTTCAAAAAGACT 377
DB 1317 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1258
QY 378 TAAAGATGATTCGAATACCTTTGAGAAATAGTCTATCAAAAGAAACGTGATTAAGTTGC 437
DB 1257 AAAAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1198
QY 438 GGAAGAACTTCTTTTAAGAGCCTTCGTAGCATCTTAATGATCATGTGTGAAA 497
DB 1197 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1138
QY 498 AAAAATTAAAGATTTGGCCCTGTCATGGCGAGGAGAAAGTGAAGTTAACTTGTG 557
DB 1137 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1078
QY 558 TCTGAACAGAAAGACATGTAGCAATTTTAAATTGAAAAAGATGAGCGCTAC 617
DB 1077 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1018
QY 618 TCTTAAACAGATGTTTCAGACGCTAGGAACTTTTAAAGAAACATGCTTGAATT 677
DB 1017 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 959
QY 678 ACTGACCAATGCTATTTTACATTGGAATTCGCGAGCGACGATTAATTAATGTAT 737
DB 958 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 899
QY 738 TGAATTGGAGGAAATGCGCAAGAACAAATATGCTTATATGCCACGAGCCGACTT 797
DB 898 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 839
QY 798 TGAATCACTAGCGCAGAGCTTACATTCAGAGATATAGGCGCTGGAAGCTTTATTA 857
DB 838 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 779
QY 858 GAAAGTGAAGAGATGAGAGTTTATGGAAGATCATCTAAGAGATGCGACGCTTT 917
DB 778 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 719
QY 918 GTTGCAATGTTGATCCAAAGATTTAGTCTTAAATAAAGAGCAAGAGAAATGCGA 977
DB 718 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 659
QY 978 AGAGCCCTTCAAAAAGCTGCAAAATCTCATGACATGAGCTTTAGAAAAGTTATG 1037
DB 658 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 599
QY 1038 TAAAGAAATGCTTTAATGATGAGAAAGTGAAGAAATGCAAAATGATAT 1097
DB 598 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 539
QY 1098 TAAACAACTTGCAAAATTTTCACTTCAAAAGTCACTATATCGCTTTTGTATCAAC 1157
DB 538 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 479
QY 1158 AAAAGGAAATTAATGAATTTGTTGATGGAGGCTTCCACATTTCTTAGACAGAA 1217
DB 478 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 419
QY 1218 TTGTGGAATTTGAGTCTATTTGTTTCAATTTTGAATAAATGTCAGATGAGAAA 1277
DB 418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 359
QY 1278 TGCATGTAAAAATATTAAGACAACTGTTACAAAAGAGACTTGTGACGGGCAATTA 1337
DB 358 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 299

```

```

QY 1338 AGTCTGCAAGAAATATGCGAGGAATGTTACATGTTCAACAAAGCTGCTTGA 1397
DB 298 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 239
QY 1398 GTTTCACACAGAAATAGTAAAGTATGTGAAAAGTGAAGAAAGAAACAA 1448
DB 238 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 188

RESULT 11
CL119201 1256 bp DNA linear GSS 05-JAN-2004
LOCUS ISB1-76J12_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-76J12,
DEFINITION genomic survey sequence.
ACCESSION CL119201
VERSION CL119201.1 GI:40612836
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1256)
AUTHORS Kremetzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTAATG
Class: BAC ends
High quality sequence start: 233
High quality sequence stop: 402.
Location/Qualifiers
source
1..1256
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-76J12"
/clone_1lb="ISB1"
/note="Vector: pBelBAC11, ISB-1 Xenopus tropicalis BAC
Library Segment 1"

ORIGIN
Query Match 3.6%; Score 111; DB 9; Length 1256;
Best Local Similarity 38.2%; Pred. No. 4.8e-12;
Matches 471; Conservative 0; Mismatches 763; Indels 0; Gaps 0;

QY 78 AAAGAGATGCAATGAGAGAGATGCAAAATACTAAATAATATTGCCAAGA 137
DB 23 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 82
QY 138 ATTGACTGAGCAAACTTAATATAGAACATGTCACAGAAACTTAAAGCTTTTGGCA 197
DB 83 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 142
QY 198 AGATGAAAGCAGATCAAAAATGCAAGAACTGAAGCCCAATTTTGAAAAAATGTAC 257
DB 143 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 202
QY 258 TACAATCAAGGAATCTTAAAGAGCAATTTAAATAAATTCAGATTATACGATTA 317
DB 203 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 262
QY 318 GGATTGCAAGGAATGAACAATGCTTATTTTGGAGGAGTATGTTCAAAAGAACT 377
DB 263 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 322
QY 378 TAAAGATGATTCGAATCTTTGAGAAATTAAGTGTATCAAAAGAAACGTGATTAAGTTGC 437

```


[illegible]

RESULT 12	1491 bp	DNA	linear	GSS 24-OCT-2003
CG753221/c				
LOCUS				
DEFINITION	CG753221			
	P048-5-A01.za Ppa Scort	BAC library	<i>Pristionchus pacificus</i>	genomic.c.
ACCESSION	CG753221			
VERSION	CG753221.1	GI:37977480		
KEYWORDS	GSS.			
SOURCE	<i>Pristionchus pacificus</i>			
ORGANISM	<i>Pristionchus pacificus</i>			

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
1 (bases 1 to 1491)	Eukaryotes: Metazoa: Nematoda: Chromadorea: Diplogasterida: Neodiplogasteridae; Pristionchus.					
1 (bases 1 to 1491)	Srinivasan,J., Sins,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buitjer,J., van der Meulen,M. and Sommer,R.J.	An integrated physical and genetic map of the nematode Pristionchus pacificus	Mol. Genet. Genomics 269 (5), 715-722 (2003)	22835951	12884007	
	Contact: Sommer RJ					
	Evolutionary Biology					
	Max-Planck-Institute for Developmental Biology					
	Spermannstr. 37-39, Tuebingen D-72076, Germany					
	Tel.: 00497071601371					
	Fax: 00497071601498					
	Email: ralf.sommer@tuebingen.mpg.de					
	Class: BAC ends.					
FEATURES	source	location/Qualifiers				
		1..1491				
		/organism="Pristionchus pacificus"				
		/mol_type="genomic DNA"				
		/strain="California"				
		/db_xref="taxon:54126"				
		/clone_lib="Ppa ECoRI BAC Library"				
		/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."				
ORIGIN	Query Match	3.6%; Score 111; DB 9; Length 1491;				
	Best Local Similarity	37.5%; Pred. No. 4.8e-12;				
	Matches 528; Conservative	0; Mismatches 880; Indels 0; Gaps 0;				
Qy	72	AATCTAAGAGATGAGCTTAAGTGCAGAGATGCAGAAAACTAATAATATTG	131			
Db	1432	AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1373			
Qy	132	CCAGAAATTGACTGAGCAAACTAATATAGAAACAAGTACACAGAAACTTAAGGTT	191			
Db	1372	NAANNAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1313			
Qy	192	TTGGGAAGATGGAAGACAGATACAAATATGCAAGAACTGAAAGCCATATTGGA	251			
Db	1312	ANAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1253			
Qy	252	ATGACTACATCAAGAGAACTTAAGAAGCAATTAAATAAATAAATTCAGATTATAC	311			
Db	1252	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1193			
Qy	312	GGATAGAGTTGCAAGAGATGAAACAATGCTTTTGGAGGAGTAGTTCAA	371			
Db	1192	AAANNAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1133			
Qy	372	AGAACTTAAGATGATTCATCTTGAATAATAGTCTATCAAAAGAAAGTGTAA	431			
Db	1132	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1073			
Qy	432	AGTGGCGAAGAGTCTTTTAAGAGACCTTCGAGGATCTTAAATGATCAGTCATATG	491			
Db	1072	AAANNAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1013			
Qy	492	TGAAAAAACTTAAGAGATTGCTTCATGCGGAGGAAAGTAGATTAACAA	551			
Db	1012	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	953			
Qy	552	CTTGTGCTGACCAAGAGACATGTAAGATATTTTAATGAAAAAGTAAGAGTG	611			
Db	952	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	893			
Qy	612	CGTACTCTTAAGACAGATGTTTCAGAGACCTAGAGAGTTTAAAGAAAGCATGCT	671			
Db	892	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	833			

QY 672 TGAATTACTCGAACAATGCTATTTTACATTGGAAATTCGGAGACGACATATATTAA 731
 DB 832 AA 773
 QY 732 ATGTATGTAATTTGGAGGAAATGCAAGAACAAATATTTGCTTATATGCCACGAGCC 791
 DB 772 AA 713
 QY 792 CGATTTTATCCACTAGCCGAGAGCTACATATACAGAGATATAGGGCTTGAAGACTT 851
 DB 712 AA 653
 QY 852 TTATTAAGAGGTAGAGAGATGAGCTTTTATTTGAAAAGATCTCTTAAGATGCCAC 911
 DB 652 AA 593
 QY 912 AGCTTGTGGCATGTTGATCCAGATTTCTAGTCTTAATAAAAAAAAAAGACAAAGAGA 971
 DB 592 NNN 533
 QY 972 ATGCGAAGAGCCCTTCAAAAAGCTGCAAAAATCTCATGACATGAGCTTTAGAAAG 1031
 DB 532 AA 473
 QY 1032 TTTATGTAAAGAAATGCTTTATGATGATGAAAGCAAAAATGCTGAAGATTCGAAA 1091
 DB 472 NAAA 413
 QY 1092 TGATATTAACAAACTGCAAAATTTTCACTTCAAAAAGCTCACTAATACTGCTTTTGA 1151
 DB 412 AA 353
 QY 1152 TCCAAACAAAGGAAATTAATGAATTTGATGGAGAGGGTCCACATTTTACGAA 1211
 DB 352 AA 293
 QY 1212 CGAAGTTGTGGAATTTGAGATCCTATTTGTTTATTTTGAATAAATGTCAGATGG 1271
 DB 292 NAAA 233
 QY 1272 AGAAATGCAATGTAATAATATATAGAGCAATGTTACAAAGAGACTTGATGACGGGC 1331
 DB 232 AA 173
 QY 1332 AAATTAAGTCTGCAAGAAATATATGCGAAGATGTTACATGCTTCAACAAAGCTGCT 1391
 DB 172 AA 113
 QY 1392 TGAATAAGTTTCAACAGAAATTTATGTAAGATATGAGAACTGAAAAGAAAGCAAG 1451
 DB 112 AA 53
 QY 1452 AAGTTTCTCAACGATGATTTATT 1479
 DB 52 TCCTATCATTAATGTACTGATGTATT 25

RESULT 13
 AG435185/c 1843 bp DNA linear GSS 03-JUN-2004
 LOCUS Mus musculus molossinus DNA, clone:MSMg01-313C22.T7, genomic survey
 DEFINITION
 ACCESSION AG435185 GI:48078248
 VERSION AG435185.1
 KEYWORDS GSS.
 SOURCE Mus musculus molossinus
 ORGANISM Mus musculus molossinus
 REFERENCE 1 Hatcori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
 AUTHORS Hatcori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
 TITLE BAC end Sequences of Library MSMg01

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1843)
 AUTHORS Hatcori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (17-NOV-2003) Masahira Hatcori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suenhiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: hatcori@sc.riken.jp, URL: http://hnp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)

COMMENT

Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rc.riken.jp).
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyada, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@rc.riken.jp
 PRIMERS

Sequencing : T7
 LIBRARY : pBACe3.6
 Vector : pBACe3.6
 R.site 1 : EcoRI
 R.site 2 : EcoRI.
 FEATURES
 source Location/Qualifiers
 1..1843
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /sub_species="molossinus"
 /db_xref="taxon:57486"
 /clone="MSMg01-313C22.T7"
 /sex="male"
 /tissue type="mixture of kidney and spleen"
 /clone_id="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 3.6%; Score 110.2; DB 9; Length 1843;
 Best Local Similarity 31.1%; Pred. No. 7.2e-12;
 Matches 480; Conservative 0; Mismatches 1063; Indels 1; Gaps 1;
 QY 50 ATGAGAACATGTTTAACTTAATTTCTAAAGAGATGACTAAGTACGAGAGAAATGCA 109
 DB 1722 ANNN 1663
 QY 110 AAAAAAAAACTAAATAATTTGCCAAGATTTGACTGAAGCAAACTAATTTAGACAG 169
 DB 1662 NNN 1603
 QY 170 TACACAGAAACTTAAAGTTTTCGAGAGTGAAGCAAGACATACAAATGCAAGAC 229
 DB 1602 ANNN 1543
 QY 230 TGAAGCCATATTTAGAGAAAATGTACTACATCAAGGAAAATTAAAGACCAATTA 289
 DB 1542 AAAAAAAAAAAAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1483
 QY 290 AAAAAAAAAATT-CGATTTATACGATAGAGATTGCAAGAGAAATGAACACATGCTA 348
 DB 1482 AANN 1423
 QY 349 TTTTGGAGGAGTATGTTCAAAAGAACTTAAAGATGTTGCAATTACTTTGAGAAATAG 408
 DB 1422 ANNN 1363
 QY 409 TGTATCAAAAGAAACGTATAAAGTTGGCGAGAGACTTTTAAAGACCTTGTAGC 468
 DB 1362 ANNN 1303
 QY 469 GATCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 528
 DB 1302 AA 1243
 QY 529 AGGAGAAAGTATGATTAACAACTTGTGTCGAGACGAGAGACATGTAAGATTT 588
 DB 1242 AA 1183

QY 589 TTAATTGAAAAAGATTAAGAGTGCCTCTTAATAACAGATGTTTCAGCAGACTAGGA 648
DB 1182 NAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAAN 1123
QY 649 AGTTTAAAAAAGAAACATGCTCTGTAATCTCGAACATCTATTTTACATTGGAAT 708
DB 1122 ANANANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 1063
QY 709 TGGGAGCGACGATATAATTATTAATTGGAGGAAATGCGAGAACAAAT 768
DB 1062 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 1003
QY 769 ATTGCTTATATCCACAGACCCGATTTTGATCCAGTACGAGGCTACATAGCA 828
DB 1002 NAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 943
QY 829 GAGATATAGGCGTGAAGAGTTTATATAGAGTAGAGAGATGAGTTTATTTGA 888
DB 942 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 883
QY 889 AAGAAATCTTAAGAGATGCGACGCTTTGTTGATTTGATCCAGATTTCTAGCTT 948
DB 882 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 823
QY 949 AAAAAAAGACGACAAAGAGAAATGCGAGAGCCCTTCAAAAAAGCTGCAAAATCTT 1008
DB 822 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 763
QY 1009 CATGAACATGAGGCTTTAGAAAGTTTATATAGAGAAATGCTTAAATGATGAGAACG 1068
DB 762 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 703
QY 1069 AAAAAATGTGAAGATTGCAAAATGATTTAACTTCAAAATTTTCACTTCAAA 1128
DB 702 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 643
QY 1129 GTCACTAATAATCGCTTTTGTATCAAAAGAAATTAATGAATTTGATGAGGA 1188
DB 642 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 583
QY 1189 GGGTTCACACATTTCTTGAACAGAGATTGTGGAATTTGAGCTTATTTCTAT 1248
DB 582 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 523
QY 1249 TTTGAAAAAATATGTCAGATGAGAAATGATTAATAATTAAGGCAACATGTTAC 1308
DB 522 TTTAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 463
QY 1309 AAAAAAGGACTTGAATGACGGGCAAAATAAGTGTGCAAGAAATTAATGCGAGATGTTA 1368
DB 462 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 403
QY 1369 CATGATTCAAACAAAGCTGCTGAAAGTTTCAACAGAGATTAATGAATGATGAG 1428
DB 402 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 343
QY 1429 AAATCTGAAAAAGAAACAAAGAGTTTCTCAACGATGAATTTATTTCTGTGCTA 1488
DB 342 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 283
QY 1489 CAGCAGCAAAAGACGCCGCTGCTTACACATGATCTTCGATGAATTAATCTTTTA 1548
DB 282 NAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 223
QY 1549 CGACAACTGATCAAAAGCGAGATTTCCCGACAGATTAATA 1592
DB 222 NAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAA 179

RESULT 14
AM33354/C
LOCUS AM33354 699 bp mRNA linear EST 31-JAN-2000

DEFINITION S20F5 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
ACCESSION AM33354
VERSION AM33354.1 GI:6829711
KEYWORDS EST.
SOURCE Pneumocystis carinii
ORGANISM Pneumocystis carinii
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 699)
Smilian A.G., Arnold J., Weise M., Wunderlich J., Staben C.,
Edman J.C., Kovacs J. and Cushion M.
Expressed sequence tags from Pneumocystis carinii
Unpublished (2000)
JOURNAL Contact: Staben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: staben@pop.uky.edu.
FEATURES
Source location/Qualifiers
1..699
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_lib="AGS-1"
/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
P. carinii organisms (3x10⁶) from a single rat (99-1-6,
sacrilficed on 3/17/99) at Cincinnati VA facilities.
Trizol extracted RNA. Oligo dt priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/Project/pneumocystis/"

ORIGIN
Query Match 3.6%; Score 109.4; DB 2; Length 699;
Best Local Similarity 56.1%; Pred. No. 9.7e-12;
Matches 317; Conservative 0; Mismatches 201; Indels 47; Gaps 4;
QY 2551 GATGTATCCGTACACAGAAATTAAGAACTTAAGTTTGGACGCAAGATTTTGA 2610
DB 681 GATGAGGTGATACAGAGTAGAGTCAAGCATTTATGCAACGAGATAGCATTTGGA 622
QY 2611 AGATATGATGTTGAAGAAAGATGTAAATTAATGAAATTCAGATTCAGATTAAGAG 2670
DB 621 TTGATTTGGATTAAGAAAGATGTAAAGCTTTAGATTAATGATGCGTTTAAAGAG 562
QY 2671 GATTGCAAGACTTAGAAGAGATATGCAAAAGATTAATAGGCTTGTGCAATCTGAG 2730
DB 561 GATTGTCCAGATTAACAAAGCTTTGCAAAATTAAGCA---CTTATGTAACTGGA 505
QY 2731 CTTCTGAGGTGAACCGCA-----CGAACGTGACGAAAGTACA 2772
DB 504 CCATTAAAAATTAAGCTTCATCATACAGAGAAATAACAGAAACGAAAGCAAG 445
QY 2773 ACGCACTACAAACAAACAAACCGCTTGCAGATCGAAGCAAG-----GAAT 2824
DB 444 ACGGAAACAAAGACGAAACAAAGATGAGCAAGCTGATGAAAGACCGTTGAGAG 385
QY 2825 GCAATCTTACAGACACAGACATGAGTTTACAGACATCGACACACAGACAGT 2884
DB 384 ACTGTTACAGAAACCAAGTCAATCAATGAGTGAAGTCAATTCATTCAGATGAG 325
QY 2885 CTACTATCACTTACATCAATCAATCAATTAATTAATGATCAACAGAGGAGTCAAC 2944
DB 324 CAACGAGTACGTCAACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 265
QY 2945 CAACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2986
DB 264 CTACCAATGTACCAACGATTCAGAAAGAGACGAAAGAGAGATGATTAAGAG 205
QY 2987 TGAACGAGTGAAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3046

, Sun Oct 30 10:37:52 2005

us-10-654-416-13.rst

Page 15

Search completed: October 29, 2005, 08:57:50
Job time : 9463 secs

This Page Blank (uspto)

301 CAGATTATACGATTAAGGATTGCAAGAAATGAACAATGCTTATTTTGGAGGA 360
| | | | |
Db 301 CAGATTATACGATTAAGGATTGCAAGAAATGAACAATGCTTATTTTGGAGGA 360
| | | | |
Oy 361 GATATGTCAAAAAGACTTAAAGATGATTGCAATACCTTTGAGAAATTAAGTCATCAAAAG 420
| | | | |
Db 361 GATATGTCAAAAAGACTTAAAGATGATTGCAATACCTTTGAGAAATTAAGTCATCAAAAG 420
| | | | |
Oy 421 AAACGTGTAAAGTTGCGGAGAAAGTTCTTTTAAAGACACTTCGTAGCGATCTTAATGA 480
| | | | |
Db 421 AAACGTGTAAAGTTGCGGAGAAAGTTCTTTTAAAGACACTTCGTAGCGATCTTAATGA 480
| | | | |
Oy 481 TCAGTCATATGTGAAAAAACTTAAAGATTTGCCCTGTCAATGGGAGGAAAGTAT 540
| | | | |
Db 481 TCAGTCATATGTGAAAAAACTTAAAGATTTGCCCTGTCAATGGGAGGAAAGTAT 540
| | | | |
Oy 541 GAGTTAAACAACTTGTCGTGAACAGAAAGACATGTAAAGATTTTAAATTTGA 600
| | | | |
Db 541 GAGTTAAACAACTTGTCGTGAACAGAAAGACATGTAAAGATTTTAAATTTGA 600
| | | | |
Oy 601 GATAAAGATGCGGTACTCTTAAACAGATGTTTCAAGACACTAGAAAGTTTAAAAA 660
| | | | |
Db 601 GATAAAGATGCGGTACTCTTAAACAGATGTTTCAAGACACTAGAAAGTTTAAAAA 660
| | | | |
Oy 661 GAAACATGCTTGGAATTACTGSAACAATGCTATTTTACATTTGGAATTTGGGAGACAC 720
| | | | |
Db 661 GAAACATGCTTGGAATTACTGSAACAATGCTATTTTACATTTGGAATTTGGGAGACAC 720
| | | | |
Oy 721 GATATATTTAAATGTATTAATTTGGAGAAAAATGCAAGAACAAATATTTGCTTAATATG 780
| | | | |
Db 721 GATATATTTAAATGTATTAATTTGGAGAAAAATGCAAGAACAAATATTTGCTTAATATG 780
| | | | |
Oy 781 CCACACAGACCCGATTTTATCTCAACTAGCCAGAGGCTACATATAGCAGAGATTAAGG 840
| | | | |
Db 781 CCACACAGACCCGATTTTATCTCAACTAGCCAGAGGCTACATATAGCAGAGATTAAGG 840
| | | | |
Oy 841 CTGGAAGAGTTTATTAAGAAAGTAAAGAGATGAGAGATTTTATTTGGAAGATATCTA 900
| | | | |
Db 841 CTGGAAGAGTTTATTAAGAAAGTAAAGAGATGAGAGATTTTATTTGGAAGATATCTA 900
| | | | |
Oy 901 AGAGATGCGACAGCTTTGTGGCATTTGTGATCCAAGATTCATGCTTAATTTAAAAAAGAC 960
| | | | |
Db 901 AGAGATGCGACAGCTTTGTGGCATTTGTGATCCAAGATTCATGCTTAATTTAAAAAAGAC 960
| | | | |
Oy 961 GACAAAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAATCTCATGAAATGATG 1020
| | | | |
Db 961 GACAAAGAAATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAATCTCATGAAATGATG 1020
| | | | |
Oy 1021 GCTTTAGAAAGTTTATGTAAAGAAATGCTTAAAGTATGATGAAACGAAAAATGTGA 1080
| | | | |
Db 1021 GCTTTAGAAAGTTTATGTAAAGAAATGCTTAAAGTATGATGAAACGAAAAATGTGA 1080
| | | | |
Oy 1081 GAATGCAAAAATGATATTAACAAAACCTTGCAAAAATTTTCACTTCAAAAGTCACTAATAT 1140
| | | | |
Db 1081 GAATGCAAAAATGATATTAACAAAACCTTGCAAAAATTTTCACTTCAAAAGTCACTAATAT 1140
| | | | |
Oy 1141 CGTCTTTTGAATCCAAACAAAGAAATTAAGAAATTTGTGATGGAAGGTTGCCAAC 1200
| | | | |
Db 1141 CGTCTTTTGAATCCAAACAAAGAAATTAAGAAATTTGTGATGGAAGGTTGCCAAC 1200
| | | | |
Oy 1201 TTTCTTACGAAGAAATTTGCGAAATTTGAGAGTCTTATTTGTTTCTATTTTGA 1260
| | | | |
Db 1201 TTTCTTACGAAGAAATTTGCGAAATTTGAGAGTCTTATTTGTTTCTATTTTGA 1260
| | | | |
Oy 1261 TGTCCAGATGGAAGAAATGCAATGTAAATATTAAGACAACTGTACAAAAGAGACTT 1320
| | | | |
Db 1261 TGTCCAGATGGAAGAAATGCAATGTAAATATTAAGACAACTGTACAAAAGAGACTT 1320
| | | | |
Oy 1321 GATGACGGGCAAAATTAAGTGTGCAAGAAATATGCGAGAAATGTACATGTTCAAAAC 1380
| | | | |
Db 1321 GATGACGGGCAAAATTAAGTGTGCAAGAAATATGCGAGAAATGTACATGTTCAAAAC 1380
| | | | |
Oy 1381 AAAAGCTGGCTTGAAGAAAGTTTCAACAAGATTTAGTAAAGTATGTGAGAAACTGAAAAA 1440
| | | | |

1381 AAAAGCTGGCTTGAAGAAAGTTTCAACAAGATTTAGTAAAGTATGTGAGAAACTGAAAAA 1440
| | | | |
Oy 1441 GAAAAACAAAGAAATTTCTCAACAGATGAATTTATTTCTGTGTGTACAGCCAGCAAAA 1500
| | | | |
Db 1441 GAAAAACAAAGAAATTTCTCAACAGATGAATTTATTTATTTCTGTGTGTACAGCCAGCAAAA 1500
| | | | |
Oy 1501 GCAGCCCGGTTGCTTACATGATCTTCCGAATGAACCTATCTTTTAAACGAACAACCTG 1560
| | | | |
Db 1501 GCAGCCCGGTTGCTTACATGATCTTCCGAATGAACCTATCTTTTAAACGAACAACCTG 1560
| | | | |
Oy 1561 GATCAAAAAGCAATTTTCCCGACAGATTAATAATTGCAAGAAATTTGGGAGAAAGTCCAA 1620
| | | | |
Db 1561 GATCAAAAAGCAATTTTCCCGACAGATTAATAATTGCAAGAAATTTGGGAGAAAGTCCAA 1620
| | | | |
Oy 1621 GATTTAGAGAGATTTCAAAAAGAAATTAACATGCGCATGTCAATACCTGAGACAGCAATGC 1680
| | | | |
Db 1621 GATTTAGAGAGATTTCAAAAAGAAATTAACATGCGCATGTCAATACCTGAGACAGCAATGC 1680
| | | | |
Oy 1681 AATGCTTTGGGACTACAGAAATTTTAAAGCAGGTTTTTATGATGAACACAAGATATCT 1740
| | | | |
Db 1681 AATGCTTTGGGACTACAGAAATTTTAAAGCAGGTTTTTATGATGAACACAAGATATCT 1740
| | | | |
Oy 1741 TTGAAAGACCAAGAAAGTTGTGTAAATTAACCTTAAAGAAAGTGTAATAATGTGCTAGA 1800
| | | | |
Db 1741 TTGAAAGACCAAGAAAGTTGTGTAAATTAACCTTAAAGAAAGTGTAATAATGTGCTAGA 1800
| | | | |
Oy 1801 AGAGAGATGACCGTTTCTCTTTGTATGTGTCTTCCAAAACGCTACGCTGAGCTGATG 1860
| | | | |
Db 1801 AGAGAGATGACCGTTTCTCTTTGTATGTGTCTTCCAAAACGCTACGCTGAGCTGATG 1860
| | | | |
Oy 1861 GTAAAGACGTGAAGACAGGTGTGAAGTATCAAAAAAATATTAAGCTTCATATATTT 1920
| | | | |
Db 1861 GTAAAGACGTGAAGACAGGTGTGAAGTATCAAAAAAATATTAAGCTTCATATATTT 1920
| | | | |
Oy 1921 ATTGAATTTCTTGAATTAATAACAAATTAATAACAACCTGGAAGAAATTTGCCCTCT 1980
| | | | |
Db 1921 ATTGAATTTCTTGAATTAATAACAAATTAATAACAACCTGGAAGAAATTTGCCCTCT 1980
| | | | |
Oy 1981 TGGCATACGTATTCATATAGATTTTCACTAATTTGTCAGAGTCTTACGAAAGAGATATG 2040
| | | | |
Db 1981 TGGCATACGTATTCATATAGATTTTCACTAATTTGTCAGAGTCTTACGAAAGAGATATG 2040
| | | | |
Oy 2041 TGTACAAAATCAAGAAAGCATGTGAGCCGTTCTATAAAGAAAGGCTTGGAAGATGCT 2100
| | | | |
Db 2041 TGTACAAAATCAAGAAAGCATGTGAGCCGTTCTATAAAGAAAGGCTTGGAAGATGCT 2100
| | | | |
Oy 2101 CTCAAGATGAGCTTCAAGAAATTTGACTGATTAATCTAAATGTGAACCTGCAATTGAA 2160
| | | | |
Db 2101 CTCAAGATGAGCTTCAAGAAATTTGACTGATTAATCTAAATGTGAACCTGCAATTGAA 2160
| | | | |
Oy 2161 AGATATTGTACAGTAGCGGAAACGTAAATTAATGCGTCAATCAAGTGTATGCAAAAGCT 2220
| | | | |
Db 2161 AGATATTGTACAGTAGCGGAAACGTAAATTAATGCGTCAATCAAGTGTATGCAAAAGCT 2220
| | | | |
Oy 2221 AAACCAAAGATTAACCTTGGAAGAGTATGAGAGTCTTGAAGAAAGAACTCTGTGAGAA 2280
| | | | |
Db 2221 AAACCAAAGATTAACCTTGGAAGAGTATGAGAGTCTTGAAGAAAGAACTCTGTGAGAA 2280
| | | | |
Oy 2281 TTAAGTGAAGAAAGTGAAGAACATGCAAAAGCACTTAACAAGAAATTTAGCAACCGGCA 2340
| | | | |
Db 2281 TTAAGTGAAGAAAGTGAAGAACATGCAAAAGCACTTAACAAGAAATTTAGCAACCGGCA 2340
| | | | |
Oy 2341 GCTGATCTTAAAAAAGATTTATTAAGACATATGAGAACTTTAAGAAACGTGACAGAGAAACA 2400
| | | | |
Db 2341 GCTGATCTTAAAAAAGATTTATTAAGACATATGAGAACTTTAAGAAACGTGACAGAGAAACA 2400
| | | | |
Oy 2401 ATGAACAAGTCCAGTCTTTGTTGTCACTCATTAAGAAAAACGAAAGTATGTATCAAAA 2460
| | | | |
Db 2401 ATGAACAAGTCCAGTCTTTGTTGTCACTCATTAAGAAAAACGAAAGTATGTATCAAAA 2460
| | | | |
Oy 2461 AGTATATGCAAAAACAGGATTAAGAAATGCCGTTCAACGGAATTCAGATATCCACAAA 2520
| | | | |

Db 1078 AAGGATTTAGAAAGATTTTAAAAAATCATGCAAAACCTCAACCAACGATCTTTAAA 1137
Qy 1138 AATCGTCTTTTGTATCAACAAGAAAATATGAAATTTGTGATGGAAAGGTTGGCA 1197
Db 1138 AACCATCTTTATGATCC-----AATGATAAATTTGTTGATGGAGAAAATGCGCA 1188
Qy 1198 ACATTTCTTACGACGAAGATTGTGCGAAATTTGAGTCCATTTGTTTCTATTTGAAAA 1257
Db 1189 ACATTTCTTACGAAAGACTGTGCAAGATTGGATCTTATTTGTTTAACTACCAAAAA 1248
Qy 1258 AAATGTCAGATGAGAAAATGCAATGTAATAATATAGCAACATGTTACAAAGAGAG 1317
Db 1249 GCTTGTCCAAATGGCCAAAGAGATGTATGAAATTTGAGGGCAGCGTGTATAGAGAGG 1308
Qy 1318 CTTGATGACGGGCAAAATTAAGTGTGCAAGAAAATATGCGAGATGTTATCATGTTCA 1377
Db 1309 CTTGATGACGGGCAAAATTAAGTGTGCAAGAAAATATGCGAGTATTTACGTGTTCA 1368
Qy 1378 AACAAAGCTGGCTTGAAGATTCTCAACAGATTAATTAAGTATGAGAACTGAAA 1437
Db 1369 AATCAAAATTTGGCTTAAAGATTTTCAACAAAGATTAATTAAGTATGAGAGAGCT--A 1425
Qy 1438 AAAGAAAACAAGAAAGTTTCTCAACGATGAATTAATTTCTGTGTACAGCCAGCA 1497
Db 1426 AAAGAAAATTAAGAAAGTTTCCCAACGATGAATTTGTTCTGTGTACAGCCAGCA 1485
Qy 1498 AAAGCAGCCCGGTGCTTACACATGATCTTCGATGAAAACCTATCTTTTACGACAA 1557
Db 1486 AAAGTCGACGATTAATTAACACATCATCAAAATGAGGGTTATCTTTTACGACAA 1545
Qy 1558 CTGATCAAAAGCGAATTTCCGACAGATTAATAATTTGCAAGAAATTTGGGAGAAAGTGC 1617
Db 1546 TTGGATCAAAAGAGATTTTCCGACAGATTAAGATCTGCAAGAAATTTGGGAGAAAGTGC 1605
Qy 1618 CAAGATTAAGAGAGATTTCAAAAGAAATTAATGAGCAATGATCAATCACTGGAGAGCA 1677
Db 1606 CAAGATTAAGAGAGATTTCAAAAGAAATTAATGAGCAATGATCAATCACTGGAGAGCA 1665
Qy 1678 TGCAATCGCTGGGACTACAGAAATTTTAAAGCAGTTTATTTGATGACACAAAGAT 1737
Db 1666 TGCAATCGCTGGGACTACAGAAATTTTAAAGCAGTTTATTTGATGACACAAAGAT 1725
Qy 1738 ACTTGAAGAACGACAAAGATTTGTGTAATTAATCTTAAAGAAAGATGTAATATGATCT 1797
Db 1726 ACTTGAAGAACGACAAAGATTTGTGTAATTAATCTTAAAGAAAGATGTAATATGATCT 1785
Qy 1798 AGAAGAGGATGACCGTTTCTTTGTATGTGCTTCCAAAAGCTACGTTGAGGCTG 1857
Db 1786 AGAAGAGGATGACCGTTTCTTTGTATGTGCTTCCAAAAGCTACGTTGAGGCTG 1845
Qy 1858 ATGCTAAAGACGTGAAGACAGTGTGTAATTTCAAAAAAATATTAAGCTTCATAT 1917
Db 1846 ATGCTAAAGACGTGAAGACAGTGTGTAATTTCAAAAAAATATTAAGCTTCATAT 1905
Qy 1918 ATTATTAATTTCTTGAATAATTAATCAAAATTAATTAACACACTGGAAGAAATGTCCC 1977
Db 1906 ATTATTAATTTCTTGAATAATTAATCAAAATTAATTAACACACTGGAAGAAATGTCCC 1965
Qy 1978 TCTTGCAATAGTATTTGCAATGATTTTCACTTAATTTGCGAGCTTAACGAAGAGAT 2037
Db 1966 TCTTGCAATAGTATTTGCAATGATTTTCACTTAATTTGCGAGCTTAACGAAGAGAT 2025
Qy 2038 AGTTGTACAAAAATCAAGAAAGCATCGTAGCCGTTCTATTAAGAAAGAGCCTTGAAGAT 2097
Db 2026 AGTTGTACAAAAATCAAGAAAGCATTTGTAGCGGTTCTATTAAGAAAGAGCCTTGAAGAT 2085
Qy 2098 GCTTCAAAAGTAGAGCTTCAAGAAAATTTGACTGATTAATCTTAAATGTAAGCTGCATTTG 2157
Db 2086 GCTTCAAAAGTAGAGCTTCAAGAAAATTTGACTGATTAATCTTAAATGTAAGCTGCATTTG 2145
Qy 2158 AAAAGATATTTGTAAGTAGAGCGGGAACGTAATTAATGCTCAATGAGGCTTAAGCAAA 2217

Db 2146 AAAAGATATTTGTAAGTAGAGCGGGAACGTAATTAATGCTCAATCAGTGCCTTAGCAAA 2205
Qy 2218 GCTTAACCAAGATTAATCTTGAAAGAGTAGAGATGCTAGAAAGAACTCTGTGAG 2277
Db 2206 GCTTAACCAAGATTAATCTTGAAAGAGTAGAGATGCTAGAAAGAACTCTGTGAG 2265
Qy 2278 AAATTAAGTAAGAGATGGAAGAACAGTGCAGTAACCAATTAACGAATTTAGCAACCG 2337
Db 2266 AAATTAAGTAAGAGATGGAAGAACAGTGCAGTAACCAATTAACGAATTTAGCAACCG 2325
Qy 2338 GCAGCTGATCTTAAAAAAAATTTATTAAGCATATGAGAACTTTAAGAAACGTGAGAGAA 2397
Db 2326 GCAGCTGATTTTAAAAAAAATTTATTAAGCATATGAGAACTTTAAGAAACGTGAGAGAA 2385
Qy 2398 GCAATGAACAAAGTCAGTCTTTTGTGCTCACTCACTTAAGAAAAAGAAAGTATGATCA 2457
Db 2386 GCAATGAACAAAGTCAGTCTTTTGTGCTCACTCACTTAAGAAAAAGAAAGTATGATCA 2445
Qy 2458 AAAAGTAATAGCAAAAAACAGGATTAAGAAATGCGCTTTCAACGGAATTCAGATTAACA 2517
Db 2446 AAAAGTAATAGCAAAAAACAGGATTAAGAAATGCGCTTTCAACGGAATTCAGATTAACA 2505
Qy 2518 AAACATGTAATAATTAACGAGAGGAGTTAAGATGTAATCCGTAACGAATTAAGAACT 2577
Db 2506 AAACATGTAATAATTAACGAGAGAGTTAAGATGTAATCCGTAACGAATTAAGAACT 2565
Qy 2578 AAACATTTGATTTGGCAGCAAGAAATTTGGAAGATATGATATTTGAAGAAAGATGT 2637
Db 2566 AAACATTTGATTTGGCAGCAGAAATTTTGAAGATATGATATTTGAAGAAAGATGT 2625
Qy 2638 AATAAATTTGAATCGAATTTGAGAAATTTAAGAGAAATTTGCAAGAAAGTATGC 2697
Db 2626 AATAAATTTGAATCGAATTTGAGAAATTTAAGAGAAATTTGCAAGAAAGTATGC 2685
Qy 2698 AAAAAGATTAATTAAGCTTTGTCGAATCTGAAGCTCTGAGAGTGAAGCCGCAAGAAACA 2757
Db 2686 AAAAAGATTAATTAAGCTTTGTCGAATCTGAAGCTCTGAGAGTGAAGCCGCAAGAAACA 2745
Qy 2758 GTGACAGAAATGACACGACAACTTAACAACAACAACAACGTTGCCGATCCGAAGCA 2817
Db 2746 GTGACAGAAATGACACGACAACTTAACAACAACAACAACGTTGCCGATCCGAAGCA 2805
Qy 2818 ACGGAATGCAAAATCTTACACACACACACACACATGCGTTAACAAGATTCGACACACA 2877
Db 2806 ACGGAATGCAAAATCTTACACACACACACACACATGCGTTAACAAGATTCGACACACA 2865
Qy 2878 AGCAGTCTACTTACATCTTAACCATCAATCAAAAAATTAATTAATGATCAATCAAGAGCGCA 2937
Db 2866 AGCAGTCTACTTACATCTTAACCATCAATCAAAAAATTAATTAATGATCAATCAAGAGCGCA 2925
Qy 2938 TGCAAAACCAACGATGTAACGACAGAGAGAGATGATGACAGAGACGTGAACCGAGT 2997
Db 2926 TGCAAAACCAACGATGTAACGACAGG-----GGATGATGACAGAGAGGTGAACCGAGT 2979
Qy 2998 GAGGGGCTGAGAGATGAGTGGGTGGAATGTGATGAGGGGGGTGAATGTAAGCAATGGTTATT 3057
Db 2980 GAGGGGCTGAGAGATGAGTGGGTGGAATGTGATGAGGGGGGTGAATGTAAGCAATGGTTATT 3039
Qy 3058 TCGTTCAATGATTTAG 3072
Db 3040 TCGTTCAATGATTTAG 3054

RESULT 3
US-09-762-724-7
; Sequence 7, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/09/762,724

CURRENT FILING DATE: 2001-02-09
 PRIOR APPLICATION NUMBER: PCT/US99/18750
 PRIOR FILING DATE: 1999-08-17
 PRIOR APPLICATION NUMBER: US 60/096,805
 PRIOR FILING DATE: 1998-08-17
 NUMBER OF SEQ ID NOS: 26
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 7
 LENGTH: 3084
 TYPE: DNA
 ORGANISM: Pneumocystis carinii sp. f. hominis
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)..(3084)
 US-09-762-724-7

Query Match 73.7%; Score 2262.6; DB 4; Length 3084;
 Best Local Similarity 84.7%; Pred. No. 0;
 Matches 2640; Conservative 0; Mismatches 399; Indels 78; Gaps 6;

QY 1 ATGCGCGCGCGGTCAAGCGGCGAGGACACGAGTACGTTGATGAGGAACAT 60
 Db 1 ATGCGCGCGCGGTCAAGCGGCGAGGACACGAGTACGTTGATGAGGAACAT 60
 QY 61 GTTTACCTTAACTTAAAGAGATGCACTAAGTGAAGAGATGCAAAATACTA 120
 Db 61 GTTTACCTTAACTTAAAGAGATGCACTAAGTGAAGAGATGCAAAATACTA 120
 QY 121 AAAAAATATGCGAAGATTGCTGAAGCAAACTAAATATAGAACAGTACAGAAAA 180
 Db 121 GAAGATATATGCGAAGATTGCTGAAGCAAACTAAATATAGAACAGTACAGAAAA 180
 QY 181 CTTAAGGTTTTCGGAAGATGCAAGATACAAATTCGAAAGAACTGAAAGCCAT 240
 Db 181 TTTAAAGATTTCTGTGATTAAGGGAACGAATGAAATGTCAAGATCTTAAAGCAAA 240
 QY 241 ATTGAGAAAAATGATCAATCAATCAAGAACTTAAAGAACTTAAAGAAATTT 300
 Db 241 GTCAATCAAAATGATCAATTTTCAAGAACTTCAAGAACTGCTAGAAAAAATTT 300
 QY 301 CAGATTAACGATTAAGATTTGCAAGAAAGATGACAAATGCTATTTTGAAGGA 360
 Db 301 TCAGATTAACGATTAAGATTTGCAAGAAAGATGACAAATGCTATTTTGAAGGA 360
 QY 361 GTATGTTCAAAAGAACTTAAAGATGATGCAATCTTTGAGAAATTAAGTCTATCAAAAG 420
 Db 361 GCATGTTCAAAAGAACTTAAAGATGATGCAATTAATTAAGAAATTAATCTGTTATCAAAAA 420
 QY 421 AAACGTGTAAGTTGCGGAAGAGTTCTTTAAGAGCACTCGTAGCGATCTTAATGGA 480
 Db 421 GAACGTAAGTTGCGGAAGAGTTCTTTAAGAGCACTCGTAGCGATCTTAATGGA 480
 QY 481 TCACTCATATGTGAAAAAACTTAAAGAGATTTGCCCTGTCAATGCGGAGGAAAGTAT 540
 Db 481 ACAAGCATATGTGAAAAAACTTAAAGAGATTTGCCCGAAATTAAGAAAGAAAGCAT 540
 QY 541 GAGTTAACTTGTGTCTGAACGAGAAAGACATGTAAGATTTTAAATTTGAAGAAA 600
 Db 541 GAATTTAACTGAGCTTGTCTTATCAAAACCAATGCTGTAAGCTTGTAAACAAAGGA 600
 QY 601 GATAAGAGTGGGTAATCTTAAACAGATGTTTCAAGAGCACTAGAGAAAGTTTAAA--- 657
 Db 601 AAAAGTAATGTGTAATCTTGAAGAAAGAGTTGAAGAGCACTTAAGAGAAATGAATTTG 660
 QY 658 AAAGAAATATGTCTTGAATTAATCGAAACATGCTATTTTAACTTGAAGAAATTTGCGAG- 715
 Db 661 CGAAGAAATATGTCTTGAATTAATCGAAACATGCTATTTTAACTTGAAGAAATTTGCGAG- 720
 QY 716 -----ACGAGATATTAATATGT 735
 Db 721 GACAAATCAAGTCAATTAATTAATGAAGCTGCAAAAGAAATATATGTACGAGAGTGT 780
 QY 736 ATTGAATTGGAGGAAATGCGCAAGAACAAATATATGCTTAATAGCCACGAGACCGAT 795

Db 781 GAGGAATTAACAGAAAAAGTGGAAAAAGAAATATTTGTTATATATGATCCAGATCCGAT 840
 QY 796 TTTGATCAACTAGAGCCAGAGGCTACATATAGCAGAGATATAGGCTGGAAGACTTTAT 855
 Db 841 TTTGATCAACTAGAGCCAGAGGCTACATATAGCAGAGATATAGGCTGGAAGACTTTAT 900
 QY 856 AAGAAAGTAGAGAGATGAGTGTATTTTGAAGAAATCATCTAAGAGATGCGACAGCT 915
 Db 901 AAGAGGCGAGAGAGATGAGTGTATTTTGAAGAAATCATCTAAGAGATGCGACAGCT 960
 QY 916 TTGTTGCAATGTTGATCAAGATTTCTAATTTTAAAAAAGACGACAAAGAAATGC 975
 Db 961 TTGTTGCAATGTTGATCAAGATTTCTAATTTTAAAAAAGACGACAAAGAAATGC 1005
 QY 976 GAAGAACCTCTTCAAAAGAGTGCAGAAATCTTATGAAATGAGGCTTTAAGAAATTTA 1035
 Db 1006 ATAAAGCCCTTAAAGAAAGTGCAGAAATCTTATGAAATGAGGCTTTAAGAAATTTA 1065
 QY 1036 TGTAAAGAAATGAGTTAAGTAATGATGGAACGAAAGAAATGTGAAGATTCAGAAATGAT 1095
 Db 1066 TGTAAAGAAATGAGTTAAGTAATGATGGAACGAAAGAAATGTGAAGATTCAGAAATGAT 1125
 QY 1096 ATTAACAAACTTGCAGAAATTTTCACTTCAAAAGTCACTAATAATCGTCTTTTGAATCA 1155
 Db 1126 GTTAACAAACTTGTACAGCTTACATCAATCTTAAAGAAATCGTCTTACATTTCA 1185
 QY 1156 ACAAAGAAATATGAAATTTGTTGATGGAAGAGGTTGCCACATTTCTTGAACAGAA 1215
 Db 1186 CC-----TGATGAAATTTGCGGAATGGGGAATTAACGACATTTCTTGAAGTGA 1236
 QY 1216 GATTGCGAAATTTGAGTCTTATTTGTTCAATTTTGAAGAAATGCGCAGATGAGAA 1275
 Db 1237 GATTGCGAAATTTGAGTCTTATTTGTTCAATTTTGAAGAAATGCGCAGATGAGAA 1296
 QY 1276 AATGATGTAATAATATTAAGAGCAACATGTTACAAAGAGACTTGAATGACGCGCAAT 1335
 Db 1297 GAAGCTGTATGATGATGAGGCGAGCGTTTACAAAGAGAGGCTTGAATGACGCGCAAC 1356
 QY 1336 AAGTGTGCAAGAAATATGCGAGGAATGTTACATGTTTCAACAAAGCTGCTTGA 1395
 Db 1357 AGTGTGTTGCAAGAAATATGCGAGGATTTATGCTGTTCAATCAAAAGTGTGCTTAAG 1416
 QY 1396 AAGTTGAACAGAAATTAAGTAAGTAATGAGAACTGAGAAAGAAAGAAAGAAAGT 1455
 Db 1417 GAGTTTCAACAGATTAAGTAAGTAATGAGAACTGAGAAAGAAAGAAAGAAAGT 1473
 QY 1456 TTCTCAACAGATTAATTTATTTCTGTGTGTAACGCGAGCAAGAAAGAGCGGTTGCTT 1515
 Db 1474 TTCCCAACAGATTAATTTATTTCTGTGTGTAACGCGAGCAAGAAAGCGATTTACTT 1533
 QY 1516 ACACATATCTTGAATGAAAACTATCTTTTACGACAACTGATCAAAAGCGAGAT 1575
 Db 1534 ACACACATATCAAAATGAGGTTTACCTTTTACGACAACTGATCAAAAGCGAGAT 1593
 QY 1576 TTCCCGACAGTAAATTTGCAAGGAATTTGGGGAAGAAAGTCCAAAGATTTAAGAGAGAT 1635
 Db 1594 TTTCGACAGTAAAGCTGCAAGGAATTTGGGGAAGAAAGTCCAAAGATTTAAGAGAGAT 1653
 QY 1636 TCAAGAAATTAACATGCGCATGTCATCACTGAGAGCAATGCAATCGCTTGGGAGCT 1695
 Db 1654 TCAAGAAATTAACATGCGCATGTCATCACTGAGAGCAATGCAATCGCTTGGGAGCT 1713
 QY 1696 ACAGAAATTTAAAGCAGGTTTATTTGAGTGAACAAAGATTTCTTGAAGAACCAAGAA 1755
 Db 1714 ACAGAAATTTAAAGCAGGTTTATTTGAGTGAACAAAGATTTCTTGAAGAACCAAGAA 1773
 QY 1756 AGTTGTGTAATAATCTTAAGAAAGAAAGTGAATTAAGTCTGAAGAGGAGATGAGCGT 1815
 Db 1774 AGTTGTGTAATAATCTTAAGAAAGAAAGTGAATTAAGTCTGAAGAGGAGATGAGCGT 1833
 QY 1816 TTCTCTTTGATGTTCTTCAAAAGCTTACGTTGAGCTGATGTTAAAGACGTGAA 1875

Db	1834	TTCTCTTTTGTATGTGTTTCCAAAACCGCTACGTCGATGCTGTAAGGTGTAAGACGTGAAA	1893
Qy	1876	GACAGGCTGTAAGTATTCAAAATAATTAAGCTTCATATTAATGAATTTCTTGA	1935
Db	1894	GACAGGTGTGAAGTATTCAAAATAATTAAGCTTCATATTAATGAATTTCTTGA	1953
Qy	1936	AATATAACAATAAATAACAACACTGGAAAGAAATTCCTCTTGGCATAACGATATTC	1995
Db	1954	AATATAACAATAAATAACAACACTGGAAAGAAATTCCTCTTGGCATAACGATATTC	2013
Qy	1996	AATAGATTTTCACTTAATTTGTCCAGGCTTTACGAAAGAAATGTTGTACAAAATCAAG	2055
Db	2014	AATAGATTTTCACTTAATTTGTCCAGGCTTTACGAAAGAAATGTTGTACAAAATCAAG	2073
Qy	2056	AAGCATCTGAGCCGCTTATTAATAAGAAAGCCCTTGGAAAGATGCTCTCAAGTGAAGCTT	2115
Db	2074	AAGCATCTGAGCCGCTTATTAATAAGAAAGCCCTTGGAAAGATGCTCTCAAGTGAAGCTT	2133
Qy	2116	CAAGGAAAATTTGACGTATTAATCTTAATGTGAACCTGCAATGGAAAAGATTTTGAACGTA	2175
Db	2134	CAAGGAAAATTTGACGTATTAATCTTAATGTGAACCTGCAATGGAAAAGATTTTGAACGTA	2193
Qy	2176	GCGGGAACGTAAATATGCGTCAATCAGTGGCTTATGCAAGCTAACCAAGGATTAAC	2235
Db	2194	GCGGGAACGTAAATATGCGTCAATCAGTGGCTTATGCAAGCTTAACCAAGGATTAAC	2253
Qy	2236	TCTGGAAAAGATGATGAGAGATGCTTGAAAGGAACCTGTGAGAAATTAAGTAAAGAGTG	2295
Db	2254	TCTGGAAAAGATGATGAGAGATGCTTGAAAGGAACCTGTGAGAAATTAAGTAAAGAGTG	2313
Qy	2296	GAAAGAACGTGCAAAAGCATTAACCAACAAATTAGACAAACCGGACGCTGATCTAAAAAAA	2355
Db	2314	GAAAGAACGTGCAAAAGCATTAACCAACAAATTAGACAAACCGGACGCTGATCTAAAAAAA	2373
Qy	2356	GATTATTAAGACATATGAGAACTTAAGAAACGTGCAGAGGAAGCAATGAACAAGTCCAGT	2415
Db	2374	GATTATTAAGACATATGAGAACTTAAGAAACGTGCAGAGGAAGCAATGAACAAGTCCAGT	2433
Qy	2416	CTTGTGTTTGTCACTCATTTAAGAAAAAGAAAGTAATGATATCAAAAAGTAATGCAAAAAC	2475
Db	2434	CTTGTGTTTGTCACTCATTTAAGAAAAAGAAAGTAATGATATCAAAAAGTAATGCAAAAAC	2493
Qy	2476	AAGGTAATGAATGCCGTTTCAAAACGGACTTCAAGATACCAAAAACATGTGAAAATACTA	2535
Db	2494	AAGGTAATGAATGCCGTTTCAAAACGGACTTCAAGATACCAAAAACATGTGAAAATACTA	2553
Qy	2536	CGAGAGGAGATTAAAGATGTATCCCTTAACAGATTTAGAAAGCTTAAAGACTTTGATTTGGCA	2595
Db	2554	CGAGAGGAGATTAAAGATGTATCCCTTAACAGATTTAGAAAGCTTAAAGACTTTGATTTGGCA	2613
Qy	2596	GCAGAGATTTTGGAGAGATGTAGATTTGAAGGAAGATGTATTAATTGGATCAGAT	2655
Db	2614	GCAGAGATTTTGGAGAGATGTAGATTTGAAGGAAGATGTATTAATTGGATCAGAT	2673
Qy	2656	TGCAAGATTTAAGAGAGATTGCAAGACCTTAAGAAAGATATCAAAAATAATTAAAGGCT	2715
Db	2674	TGCAAGATTTAAGAGAGATTGCAAGACCTTAAGAAAGATATCAAAAATAATTAAAGGCT	2733
Qy	2716	TGTGCGCAATCTGAAGCCTCTGGAGGTGAAACCGCACGAAACAGTGAACGAAGAATTCCTTA	2757
Db	2734	TGTGCGCAATCTGAAGCCTCTGGAGGTGAAACCGCACGAAACAGTGAACGAAGAATTCCTTA	2793
Qy	2776	ACAACTTACAAACAACAACAACCGTTGCCGATCCGAAGGCCAAGGAAATGCAAAATCCTTA	2833
Db	2794	ACAACTTACAAACAACAACAACCGTTGCCGATCCGAAGGCCAAGGAAATGCAAAATCCTTA	2853
Qy	2836	CAGACAAACAGACACATGGGTTACACAGACATGACACACACAAGACGTTACTATTCACA	2895
Db	2854	CAGACAAACAGACACATGGGTTACACAGACATGACACACACAAGACGTTACTATTCACA	2913
Qy	2896	CTTACCAATCAATCAAAAATAACAATTGACATCAACGAAGGCCATGCAAAACCAACCAAGGT	2955
Db	2914	CTTACCAATCAATCAAAAATAACAATTGACATCAACGAAGGCCATGCAAAACCAACCAAGGT	2973

```

Oy      2956  ACGCAGAGAGGAGGAATGATGTGACAGAGACGTGAAACCGAGTGAAGGCTGAGATGAGT 3015
Db      2974  ACGACACAG-----GAGTATGACAGAGACGTGAAGCCAACTGAAGCTTGAAGGCTGAGC 3027

Oy      3016  GGGTGAATGTGATGACGGGGGTGATAGTGAACATGGTTATTTGCTTCATGATTTAG 3072
Db      3028  GGGTGAATGTGATGACGGGGGTGATAGTGAACATGGTTATTTGCTTCATGATTTAG 3084

RESULT 4
US-09-762-724-5
; Sequence 5, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; TITLE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/09/762,724
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 3090
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3090)
US-09-762-724-5

Query Match 61.9%; Score 1900.6; DB 4; Length 3090;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 609; Indels 84; Gaps 8;

Oy      1  ATGGCGGGGGCGGTCAACCGGCAGACGAGGACACAGAAATACCATTTGATGAGGAACAT 60
Db      1  ATGGCGGGGGCGGTCAACCGGCAGAAAGGTGCACAGAAATACCATTTGATGAGGAGCAT 60

Oy      61  GTTTTACCTTTAATTTCTAAAGGAATGSACTAATGAGCAGCAGATGCAGAAAAAACTA 120
Db      61  GTTTTACCTTTGATTTTAAAAAAAATGATTAAGAATTCAAAATGCAAAATGCAAACTAAGTTG 120

Oy      121  AAAAAAATATTCGCAAGATTCGCTGAAGCAAAACTAAATATATGAACAATGACAGAAAA 180
Db      121  GAAATAATATTCGCAAAACATTTAAACAAATGCAAGATTTAATCCAGAAAAAAGTTACAGAAAAA 180

Oy      181  CTTAAAGTTTTCGCGAAGATGAGAAAGCGATCAAAATGCAAGAACTGAAGGCCAAT 240
Db      181  TTTAAAGATTTCTGTGTATACGGGAAACGAATGAAAAATGTCAGATCTTAAAAAACAA 240

Oy      241  ATTGAGAAAAAATGTACTACATCAAAAGSAAAACTTAAAGAAAGCAATTTAAAAAAAATTT 300
Db      241  GTCAATCAAAAATGCAATTAATTTCAAGSAAAACTTCAAAAGCTGCTGAAAAAAAATTT 300

Oy      301  CAGATTATTAACGATTAAGATTGCAAAAGAGATGAACAACAATGCTTATTTTGGAGGGA 360
Db      301  TCAGAAATTAACAGATGAGSATTGCAAAAAGATGAACAACAATGCTTATTTTGGAGGGA 360

Oy      361  GTAGTTCAAAAGAACTTAAAGAGATTCGAATCTTTGAAATTAAGTGTATCAAAAG 420
Db      361  GCAATGTCACACAGAACTTAAAGAGATCTGCAATTAATTAAGSAAATACGTGTTATCAAAA 420

Oy      421  AAAGGTGATAAGTTGGCGAAGAGATTTCTTTAAGACACTTCGTAGCATCTTAAATGGA 480
Db      421  GAAGGAAACAATGTGGCGAAGAGATTTCTTTTGAAGGCGCTTCGTGTGATCTCAATGAA 480

Oy      481  TCAGTCATATGTGAAAAAAAATTAAAGAGATTTTCCCTGTGATGAGGAGGAAAAATGAT 540

```

```
Db 481 ACAAAACATGTGAAAAAAGCTGAAAGAGTTGGCCGAAATTAAGAAAGGAAAGCGAT 540
Qy 541 GAGTTAAACAACTTGTGTCTGCAACGAGAAAGACATGTAAAGATTTTAAATGAAAAA 600
Db 541 GAAATTAACGAGGCTTTGCTTTATCAAAAAACAACTGCGTAAGTCTTTGTAACAAAAAGA 600
Qy 601 GATTAAGAAAGTCCGCTACTCTTAAACAGATGTTTACAGAGCACTAAGGAATTTTAAA--- 657
Db 601 AAAAGTAAATGTGATCTTGTGAAAGAAAGTTGAAGAGCACTTAAGAGAAATGAATTG 660
Qy 658 AAAGAAACATGTCCTGAATTAATCGAAACAATGCTAATTTTACATTGAAAAATGCGAG-- 715
Db 661 CGAGAAAAATGTCACTATTACTTGAGCAATGTACTTTTCAAGAGGAACTGTGAAGGA. 720
Qy 716 -----ACGACGATTAATTAATGT 735
Db 721 GACAAATCAAAAGTCAATAACTTAATTAATTAAGACTGCAAAAGAAATATGTACAGAGTGT 780
Qy 736 ATTGAATTTGGAGGAAATGCGCAAGAACAAATATTGCTTATATGCAACCGAGCCGAT 795
Db 781 GATGAATTAAGCAAAAGTGTGAAAGAAATATGTATTATGCAATCCAGATCCGAT 840
Qy 796 TTTGATCAACTAGGCGAGGCTTACAATAGCAGAGATATAGGCTGGAAGATTTTAT 855
Db 841 TTGATCCAACTAAGCCAGAGCTTACACTAGCAGAGCAATAGGCTGGAAGACTTTAT 900
Qy 856 AAGAAGGTAGAGAGATGAGATTTTATTGGAAGAAATCATCTTAAGAGTGCACAGCT 915
Db 901 AAGGGCCAGAGAGATGGAATTTTGTGGAAGACAACTGTTAAGAGATGCAACAGCT 960
Qy 916 TTGTGGCATTTGTGATCCAAAGATTCTAAGTCTTAAAAAAAAGACGAACAAAGAGAAATGC 975
Db 961 TTGTGGCACTACT-----TCTTAAGAAACCTTTAAAAAGAAAGAAATGT 1005
Qy 976 GAAGAAACCCCTTCAAAAAAGCTGCAAAAAATCCTCATGAAACATGAGGCTTTAGAAAGTTTA 1035
Db 1006 ATAAACCCCTTAAAAAAACCTGCAAAACCTCATGAACATGAGGCTTTAGAAATCTA 1065
Qy 1036 TGTAAAGAAAAATGCTTTAAGTAATGATGAACGAAAAAATGTGAAGATTCGAAATGAT 1095
Db 1066 TGTAAAGAAAAATAAACCAAGATGATGGAACGAAAAAATGTGATGAACCTAGAAAAAGAT 1125
Qy 1096 ATTAAACAAACTTGCAGAAATTTTCACTTCAAAAGCACTAATATGCTTTTGAATCCA 1155
Db 1126 GTTAAACAAACTTGTACAGTCTTACATCAACAAATCTTAAACACGCTTTTACATTTCA 1185
Qy 1156 ACAAAAGAAATATGAATTTGTGATGAGGAGGTTGCCAACATTTCTTACGACAGAA 1215
Db 1186 CC-----TGATGGAATTTGGCGAATGGGGAATTAACCGACATTTCTTAAGTGAAGA 1236
Qy 1216 GATTGTGCGAAATTTGAGTCTTATTTGTTTCTAATTTTGAAAAAAATGTCAGATGAGAA 1275
Db 1237 GATTGTGCGAAATCTGAATCTTATTTGCTTTTATTAAGAAACCTTGTCCAGATGTCAA 1296
Qy 1276 AATGCAATGTAATAATTAAGCAACATGTTACAAAAGAGGACTTGATGACGCGGCAAT 1335
Db 1297 GAACCTGTGATGAATGTGAGGCGAGCGTGTATTAAGAGGGCTTGTGATGACGCGGCAAC 1356
Qy 1336 AAAGTGTGCAAGAAATATGCGAGAAATGTTACATGTTTCAACAAAGCTGTGCTGAA 1395
Db 1357 AGTGTGTGCAAAAAAATATGCGAGGTTATTTGATGCTCAAAATTAAGATTTGCTTAAG 1416
Qy 1396 AAGTTTCAACAAATTAATTAAGTAAAGTATGTAGAAATGAAAAAAGAAACAAAGAAAT 1455
Db 1417 AAATTTCAACAAAGATTAAGCAAAAGTATGTAGAAACCTG---AAAGGAAATTAAGGAAGT 1473
Qy 1456 TTCTCAACAGATGAATTAATTAATCTGTGTGATACAGCCAGCAAAAGCAGCCGCTGCTT 1515
Db 1474 TTCTCAACAGATGAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1533
Qy 1516 ACACATGATCTTTCGAATGAAAACTATCTTTTACGACAAACAACTGATCAAAAGCGAGAT 1575
Db 1534 ACACATCACCATCAAAATGAGAGTTATCTTTTACGACAAACAACTGATCAAAAGAGAT 1593
Qy 1576 TTCCCGCAGATTAATAATTTGCAAGAAATTTGGGGAAGAAAGTCCCAAGATTTAGAGAGAT 1635
Db 1594 TTTCGACAGATTAAGACTGCAAGAAATTTGGGGAAGAAATTTGCCAAGATTTTAGAAAGAT 1653
Qy 1636 TCAAAAGAAATTAATGATGCCATGTCAATACAGAGCAAGCAATGCAATGCTTTGGGACT 1695
Db 1654 TCAAAAGAAATTAATGATGCCATGTCAATACAGAGCAAGCAATGCAATGCTTTGGGACT 1713
Qy 1696 ACAGAAATTTTAAAGCAGGTTTATTTGATGAACAAAGATCTTTGAAAAGCAAGAA 1755
Db 1714 ACAGAAATTTTAAAGCAGATTTTATTTGATGAACAAAGATCTTTGAAAAGCAAGAA 1773
Qy 1756 AGTTGTGTAATAATCCATAAAGAAAGATGTAAATTAATGATCTGAAAGAGAGATGACCGT 1815
Db 1774 AACTGTCAAAATTAATTTAAAGAAAGAAATGCAATTAATGATCTGAAAGAGGATGATGCT 1833
Qy 1816 TTCTCTTTGTATGTGCTTTCAAAACGCTAAGTGTGAGCTGATGATTAAGACGTGAA 1875
Db 1834 TTTTCTTTGTATGTGTTTCCAAAACGCTAATGATGAGCTGATGATTAAGACGTGCA 1893
Qy 1876 GACAGGTGAGATGATTCAAAAAATATTAAGCTTCATATATTTGAATTTCTTGAA 1935
Db 1894 GATAGGTGCAAAATATTCGAAAGAAATATGCAAGCATCAATATTAATGATTCCTTAAA 1953
Qy 1936 AATAATCAATATAATAATTAACAACCTGGAAGAAATTTGCTCCCTGTCATACGATTTGC 1995
Db 1954 AAATATCAATTAATAAGCAAGATGACGCAATATTTGTCCCTCATGACATCCATCTGC 2013
Qy 1996 AATGATTTTCACTTAATTTGTCAGTCTTACGAAAGAAATGCT---TGTACAAAATC 2052
Db 2014 GATGATTTTATCCCATTTGCTCGATCTTAAGAAAGAAATCTTCTGTCAAAATCTT 2073
Qy 2053 AAGAAGCATGTGAGCCGTTCTATTAAGAAAGAGCCCTTGAAGATGCTTCAAGTGAAG 2112
Db 2074 AAAAAATATTGCGAACATTTCAAAAAAGAAAGTTTGAAGATGCTTAAAGTGAAG 2133
Qy 2113 CTTCAGAAATTTGATGATTAATCTAATGTGAACCTGATGAAAGATATTGTACA 2172
Db 2134 CTTCAGAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2193
Qy 2173 GTAGCGGAAACGTAATTAATGCTCAATCAATGAGCTTTATGCAAAAGCTTAACCAAGAGAT 2232
Db 2194 GTATTTGAAGACGTAATTAATGCTCAATCAATGAGCTTTATGTAAGATTAATTCGAAAGT 2253
Qy 2233 AACTGTGAAG--AGTATGAGATGCTTAAGAAAGAACTGTGTGAAGATTTGTGA 2289
Db 2254 AAAAATTAAGAGCGCAATTAATAAATGTTAAGAAAGAGCTTTGTCTTAATAATTTGTGA 2313
Qy 2290 GAAGTGAAGAACGTCGCAAAAGCAATTACCAAGAAATTAGGACAAACGGGCACTGATCTA 2349
Db 2314 GAGGTGAAGACGATCAAAAGTATTAACCAAGAAATTAACAGGCTGTGAAAAAGCTTA 2373
Qy 2350 AAAAAATTAATTAAGATTAAGATTAAGAACTTAAGAAAGTGTGAGAGAAAGCAATGAAGA 2409
Db 2374 AAAAAATGTTAAGATTAAGATTAAGAACTTAAGAAAGGCAAAAAAGCAATGAAGA 2433
Qy 2410 TCAGCTTGTGTTTGTCACTCATTAAGAAAAAGAAAGTAAATGATTAATAAAGTAAATGC 2469
Db 2434 TCAGCTTGTGTTTGTCACTCATTAAGAAAAAGAAAGTAAATGATTAATAAAGTAAATGC 2493
Qy 2470 AAAAAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2529
Db 2494 AAAAAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2553
Qy 2530 ATACTAGAGAGGAGTTAAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2589
Db 2554 ATACTAGAGAGGAGTTAAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2613
Qy 2590 TTGGCAGAGAGATTAATTTGAAGATTAATGATTTGAAGAAAGATTAATTAATTTGA 2649
Db 2614 TTGGCAGAGAGATTAATTTGAAGATTAATGATTTGAAGAAAGATTAATTAATTTGA 2673
```

QY	2650	TCBAAATTGCGAAATTTAAGSAGATTGCAAAAGACTTAGAAGAAAGTATGCAAAAAGATTAAAT	2709
Db	2674	TCGGATTGCGGATTTAAAGACGATTGCGATGGTTTAAAAGAAAGTGTGCGAAAGTTGAG	2733
QY	2710	AAGGCTTGTGCGAATCTGAAGCCTCTGGAAGGTGAAGCCGACGAAACAGTGACAAAAGT	2769
Db	2734	AAGCAGTGTCAAGATCTGAAGCCTCTGGAAGGTGAAGTGCATGAATTAAGTCACGAAAGC	2793
QY	2770	ACAAACGACAACTTAACAACAACAACACGTTGCCGATCCGAAAGCAAACGAAATGCAA	2829
Db	2794	ACAAACGACGACCAACAACGACAAACGACCGTACCGATCCGAAAGGCAACGAATGCAA	2853
QY	2830	TCCTTACGACAAACAGACACATGGGTTTACAGACATGACACACACACAGCAGCTTACT	2889
Db	2854	TCCTTACGAGCAACAGATACATGGGTTTACAGACATGACACACACAGCAGCTTACC	2913
QY	2890	ATCACATCTTACCATCAATCAAAAATAACATTGACATCAACGAGCGCATGCAAAACCAAC	2949
Db	2914	ATCACATCTTACCATCAATCAAAAATAACATTGACATCAACGAGCGCATGCAAAACCAAC	2973
QY	2950	AAGAGTTCGACAGGAGGAAAGATGATGCAGAGACGGAACCGAGTGAAGGAGGACTGAGG	3009
Db	2974	AAGGTTCACGACAG-----GATGTAGACAGAGACGCTGAACCGAGTGAAGGAGATTGAG	3027
QY	3010	ATGAGTGGGTGGAATGTGATGAGGGGGGTGATAGTAGCAATGTTATTTCTTTCATGATT	3069
Db	3028	ATGAGTGGGTGAGAGGTGATGAGGGGGGTGATAGTAGCAATGTTATTTCTTTCATGATT	3087
QY	3070	TAG 3072	
Db	3088	TAG 3090	

RESULT 5
US-09-762-724-9
; Sequence 9, Application US/09762724

Patent No.: 5664053
GENERAL INFORMATION:
APPLICANT: Kovacs, et al.
TITLE OF INVENTION: Identification of a region of the major surface
TITLE OF INVENTION: glycoprotein (M5G) gene of human Pneumocystis carinii
FILE REFERENCE: 4239-58054
CURRENT APPLICATION NUMBER: US/09/762,724
CURRENT FILING DATE: 2001-02-09
PRIORITY APPLICATION NUMBER: PCT/US99/18750
PRIORITY FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: US 60/096,805
PRIOR FILING DATE: 1998-08-17
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 3081
TYPE: DNA
ORGANISM: Pneumocystis carinii sp. f. hominis
FEATURES:
NAME/KEY: CDS
LOCATION: (1)..(3030)
US-09-762-724-9

Query Match	52.9%	Score 1624.8	DB 4	Length 3081
Best Local Similarity	72.8%	Pred. No. 0		
Matches 2275; Conservative	0	Mismatches 752	Indels 99	Gaps 10

QY	1	ATGGCGCGGGCGGTCAAGCGCGAG-----GCAGCAGGGGACACAGATAGCATTTGAT	51
Db	1	ATGGCGCGGGCGGTCAAGCGCGAGCGTGCATAAAGCATCGAGGGCTTAGTATATATGATGGT	60
QY	52	GAGGAACAATGTTTAGCTTTAATTCCTAAAGAAAGATGACTAAGTGAAGCAGGATGCAAA	111
Db	61	GAGGAATATCTTTGGCTTTAATTGCA---GGAAAAAATATATATGATATATGATGATCAAA	117
QY	112	AAAAAACTAAAAAAAATATTTGCCAAGATTTGACTGAAGCAAAACTAAATATATAGAACAGTA	171

Db	118	AAAGATTAGAAAAATTTGTAGACCTTAAACGATGCAGATTAAACAGAAAAAGTT	177
Qy	172	CACGAAAACCTTAAGGTTTTTGGAGATGAAAACAGATCAAAAATGCAAGAACTG	231
Db	178	CACAAAAAACTTAAGAGTTTGTGAAAAATAAAAACAGATTCAAAATGCAAGAACTG	237
Qy	232	AAAGCCAAATTTGAGAAAAAATGATCTACATCAAAAGAAAACTTAAGAAACAATTAA	291
Db	238	AAAAAAAACCTCACTAAAAATGTACTGCATCAAGAAAACTTACGAAGCAATCAA	297
Qy	292	AAAAAAATTCAGATTATACGSAATAAAGATTGCAAGAGATGAAACAACATGCTATT	351
Db	298	AAAAAAATTCAGATTATACGSAATGAAGATTGCAAAAGAAATGAACAATGCTATT	357
Db	358	TTGAGGAGAGATGTCCAGCGGAACTTAAAGATGATTGCAATCTTGAGAAATAAGTC	417
Qy	352	TTGAGGAGAGATGTTCAAAAGAACTTAAAGATGATTGCAATCTTGAGAAATAAGTC	411
Db	412	TATCAAAAAGAAAGTGTAAAGTTGGCGGAGAAAGTTCTTTTAAGAGCACTTCGATGCAT	471
Qy	418	TATCAAAAAGAAAGTGTAAAGTTGGCGGAGAAAGCTTTTAAAGAGAGTTGCGTGAAGT	477
Db	472	CTTAATGATCAGTCATGTGTAAGAAAAAACTTAAAGAGATTTCCTGTCATGGGAGG	531
Qy	478	CTTAATGATCAGTAATCTACATGTGAAGAAAGCTCAAGAGTTTGATAGAGTTAGTCA	537
Db	532	GAAAGTATGATGTAAACAACTTGTTGTCTGAACGAAAGAGACATGTAAAGATATTTTA	591
Qy	538	GAAAGTATGATGTAAACAAAGCTTTGTCTTTATCAAAAAATGACGTCAAAACATTTGTA	597
Db	592	ATTGAAAAAGTAAAGAAAGTGGCGTACTCTTAAACAATGTTTACGACGACTG----	646
Qy	598	TTAAGAAAAACAAAAAATGTAAAGTCTTTAAACAGATGTAAACGAGCACTTGAGAG	657
Db	647	-GAAGTTTAAAAAAGAAACATGCTTGAATTAATCGAAACAATGCTATTTTACATTGA	705
Qy	658	AAAGATGATTAACGAGAAATGTTTACCACTGCTTGAACATCTATTTTATAGAGGG	717
Db	706	AATTGCGAG-----ACGACGATATAATT	729
Qy	718	AATTGTGAAGTATATCAAAATGTAAATTAATCAATCGAAGCTGTATGAATATTGGCA	777
Db	730	AAATGTATTGAATTGGAGAAAAATGCCAAGACAAAAATTTGCTTATATGCCACGAGA	789
Qy	778	GTGTGTATCATTTGGACGTAAATGTGAAGAAATAGATTATTATACACATCCGGGA	837
Db	790	CCGATTTTGAATCAACTAGGSCGAGGCTACATACAGAGGATATAGGGCTGGAAGAG	849
Qy	838	TCCGATTTCAATCCAACTTAAGTCAAAAGCTTACTGACGAAAGCATAGGACTGGAAG	897
Db	850	TTTTATAAGAAAGTATAGAGAGATGAGATTTTATTTGAAAAAATCTATTAAGATGCG	909
Qy	898	CTTTATTAAGAAAGGCGCAGAGAGAGGTTCATATTGGAAGCCTCTGTAAAGATGCA	957
Db	910	ACAGCTTTGTTGGCATTGTGTGATCAAGATTCTAGTCTTAAAAAAGACGACAAAGAG	969
Qy	958	ACTGCTCTACTGGCGCTTTTGATTTCAAAATCTAGATCTTAAGAGTCAAGTGGTAAAG--	1015
Db	970	AAATGCCAAGAGCCCTTCAAAAAAGCTGCAAAAATCTCATGAACATGAGGCTTTAGAA	1029
Qy	1016	-AATGCCAAGAGCTTTAAAGATTAACGTAAAGTTAAAGTCAATGAATTTTGGGA	1074
Db	1030	AGTTATGTATGAAGAAAAATGCTTAAAGTAAATATGATGGAAGAAAAATGGAAGATTGCA	1089
Qy	1075	GATTTTGTATCAAAATATGTAGCTGCTCAAAATTAATGAAAAGTGAAGAGATTAGAG	1134
Db	1090	AATGATATTAACAAACTTGCAAAATTTTCACTTCAAAAGTCACTAAATATCGTCTTTT	1149
Qy	1135	AAGAGTTAGCAACAGTACTTAAATTTCTTTTGAAGAAAAATTAAGAAATTAACACTCTCT	1194
Db	1150	GATCAACAAAGAAATATGAATTTGTTGATGGAAGGTTGCCAATTTCTTATGC	1208

Db 1195 -----GGATCCGGAAGATCATTCATGATATTAAGTGAAGACATTTCTTAGT 1242
Qy 1210 AAGCAAGTTGTGGAAATTTGAGTCTATTTGTTTCTATTTTGAAGAAAAATGTCAGAT 1269
Db 1243 GACATATGCTGCACAGGTTAGAGTCAGACTGTTTATTTT- -AAAAGTCAGACACT 1299
Qy 1270 GGAGAAAATGATGTAAATAATATAGACCAATGTTACAAAGAGCACTTGATGACCG 1329
Db 1300 CTTGACAAAGATGTAAATATCTGAGGACATGTTATAGAGGGCTTGAGACAA 1359
Qy 1330 GCAATTAAGTCTGCAGAAAATATGCGAGAAATGTTACATGTTCAACAAAGCTGG 1389
Db 1360 GCTAATAGAGATTCGAGAAAAGATGTACGACTGTTCTATGTTCCAGGAAAAGATGG 1419
Qy 1390 CTTGAAAAGTTTCAACAGAAATTAAGTATGTGAAACTGAAAAGAAAAGAAACAA 1449
Db 1420 TTTAAGAACTATAGAAAATAATGAGAAATGTCGAACTTAAAC- ------ 1470
Qy 1450 GGAAGTTTCTCAACGATGAATTAATTTCTGTGTACAGCCAGCAAGACCCCG 1509
Db 1471 -----ACAAGCATGAGTTGTTTGTCTATGATGATCCATTAAGCAGTCA 1521
Qy 1510 TTGCTTACATGATCTTCAATGAAAATCTATCTTTTGAAGCAACATGATCAAAAG 1569
Db 1522 ATACTTCAGCTGATTCACAGCAAGCAGCTTTTTCGGAACCAATGATCAAAAG 1581
Qy 1570 CGAGATTTCCGAGATTAATAATTTGCAAGAAATGGGAGAAAGTGCAGATTTTGA 1629
Db 1582 CGAGACTTCCAGCAAGCAAGATTCAGGAAATTAAGAAAGATGAAAGCTTTAGGG 1641
Qy 1630 GAGATTTCAAAAGAAATTAACATGCGCATGTCATACACTGAGCAGCAATGCATG 1689
Db 1642 AAGGATTCATATCAGATTAATGTCGATGTCATACGCTAAACACAGTGTGATG 1701
Qy 1690 GGGACTACAGAAATTTTAAACAGGTTTATGATGACCAAGAAATGATCTTTAAAG 1749
Db 1702 GGGACTACAGAAATTTTAAACAGGTTTATGATGACCAAGAAATGATCTTTAAAG 1761
Qy 1750 CAAGAAATGTTGTAATAATCTTAAAGAAAGTAAATTAATGCTAGAGAGGAGAT 1809
Db 1762 CATAAAACTGTAAGAAATTTTAAAGAAAGTCAATTAATGCTAGAGAGGAGAT 1821
Qy 1810 GACGTTTCTTGTGTATGTCCTCCAAACGCTACGTCGATGATGTTAAAGAC 1869
Db 1822 GATGTTTCTTGTGTATGTCCTCCAAACGCTACGTCGATGATGTTAAAGAT 1881
Qy 1870 GTGAAGACAGGTGTGAATTAATAAATAATTAAGGCTCATATATTAATTAATTT 1929
Db 1882 GTGAAGACAGGTGTGAATTAATAAATAATTAAGGCTCATATATTAATTAATTT 1941
Qy 1930 CTTGAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1989
Db 1942 CTTGAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2001
Qy 1990 TATGCAATAGATTTTCACTAATTTGTCAGAGTCTTACGAAAGAAATAGT- -TGTACA 2046
Db 2002 TATGCAATAGATTTTCACTAATTTGTCAGAGTCTTACGAAAGAAATAGT- -TGTACA 2061
Qy 2047 AAAATCAAGAGATGTCAGCCGTTCTATTAAGAAAGAGGCTTGAAGATGCTCAAA 2106
Db 2062 AATCTTAAAAATATGTCGAACTATTCACAAAGAAAGGTTTGAAGATGCTCTTAAA 2121
Qy 2107 GTAGAGTTCAAGAAAATTAATCTAATTAATGTAACCTGCTTGAAGAAATAT 2166
Db 2122 GTAGAGTTCAAGAAAATTAATCTAATTAATGTAACCTGCTTGAAGAAATAT 2181
Qy 2167 TGTACAGTACGAGGAAAGTAAATATGTCGATCAATGATGCTTACGAAAGTCAAC 2226
Db 2182 TGTACAGTATTTGAAGAAATGTAAGATTCATCAACAGTTTATGTAAGAAATATCC 2241
Qy 2227 AAGATTAATCTGTGAAGAGTATGAGATCTTGAAGAAAGTCTGTGAGAAATTAATG 2286
Db 2242 GAAAGTAAATCTAAAGAGAGATTAATGAGTGAAGAAAGCTTGTCTAAATTAATG 2301

Qy 2287 AAGAGTGAAGAAAGATGCAAGATTAACACAGAAATTAAGCAACCGGAGCTGAT 2346
Db 2302 GAGAGGTGGAACGCAATGTAAATGTTACAGCAAAATTTGAGCATGAGAAAAGAC 2361
Qy 2347 CTAAAAAAGATTAATGAATATGAGAACTTAAGAAAGTGTGAGAGAAATGATAC 2406
Db 2362 CTAAAAAAGATTTTGAAGATTTGAAGAACTTAAAAAAGGCGAGAAAGCAATGAT 2421
Qy 2407 AAGTCCAGTCTGTTTGTGATCATTAAGAAAAGCAAGTATGATCAAAAGTAT 2466
Db 2422 AATTCATCTGTTTATCATTCGTTAAGAAAGATGAAATTAATCATGAAAATGCT 2481
Qy 2467 AGCAAAAACAGATTAAGATGCGCTTCAACGCACTTAAGATTAACACAAACATG 2526
Db 2482 AGCAAAACAGATTAAGATGCGCTTCAACGCACTTAAGATTAACACAAACATG 2541
Qy 2527 AAAATCTACGAGGAGTTAAGATGATCCGTAACAGATTAAGAACTTAAGATTT 2586
Db 2542 AAAATCTACGAGGAGTTAAGATGATCCGTAACAGATTAAGAACTTAAGATTT 2601
Qy 2587 GATTTGCAAGAGATTAATGAAATGATGATGATGATGATGATGATGATGATGATG 2646
Db 2602 GATTTGCAAGAGATTAATGAAATGATGATGATGATGATGATGATGATGATGATG 2661
Qy 2647 GAATCAGATTCAGAAATTAAGAGATTTGCAAAAGCTTAAGAAAGTATGCAAAAGAT 2706
Db 2662 GAATCAGATTCAGAAATTAAGAGATTTGCAAAAGCTTAAGAAAGTATGCAAAAGAT 2721
Qy 2707 AATTAAGCTTGTGCAATCTGAAGCTTGTGAGGTGAAGCCGACGAAACAGTACAG 2766
Db 2722 CAAGAGATGATGTTGAATTAATTAACCACTGAAGATGAAGCCGACGAAACAGTACAG 2781
Qy 2767 AGTCAACGACACTACAAACAAACAAACCAACCGTTGCGCATCCGAGGCAACGGAATGC 2826
Db 2782 AGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2841
Qy 2827 AATCTTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2886
Db 2842 AATCTTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2901
Qy 2887 ACTATCATCTTACATCATCATCAAAATTAATTAATTAATTAATTAATTAATTAAT 2946
Db 2902 ACCATCATCTTACATCATCATCAAAATTAATTAATTAATTAATTAATTAATTAAT 2961
Qy 2947 ACCAAGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 3006
Db 2962 ACCAAGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 3015
Qy 3007 AGATGAGTGGGTGAAATGATGATGAGGGGGGTATGATGATGATGATGATGATGATG 3066
Db 3016 AAGATGAGTGGGTGAAATGATGATGAGGGGGGTATGATGATGATGATGATGATG 3075
Qy 3067 ATTTAG 3072
Db 3076 ATTTAG 3081

RESULT 6
US-09-762-724-1
; Sequence 1, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/09/762, 724
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096, 805
; PRIOR FILING DATE: 1998-08-17

QY 1964 AAGAAATGTCCTTGGCATAGTATTGCAATAGATTTTCACTAATTGTCAGATC 2023
DB 1964 GAGATATTTGCTTATTTGGCACCACATGCTTATTTGGACCCCAATTTGCCGATC 2023
QY 2024 TTACGAAAGAAATAGTGTACAAATAATCAAGAACGTCGTCCTTATTAAGAA 2083
DB 2024 TTGAAAAA---AATTAATGTGAAGACTTTGAAAAAATTTGCAAACTTATTAAGCAA 2080
QY 2084 AGGCTTTGGAAGATGCTCTCAAGTAGAGCTTCAAGAAAAATTAATGATTAATCTAAAT 2143
DB 2081 GAGACCTTGAATAAGCACTTATTTGATTGATGAGACATCTTGATTAAGAAAAAACT 2140
QY 2144 GTGAACCTGCATTAAGAAATATTTGTCAGTAGGGGAAAGTAATTAATGCTGCAATCA 2203
DB 2141 GCAAAACAAATCTTGATTAAGTACTGTACCTATGSGATCAACAGAAATTAACACTTA 2200
QY 2204 GTGCTTATGCAAAAGCTAACCAAGATTAATCTGAAAGATGATGAGATGCTTGA 2263
DB 2201 AAGGTTTTTGA-----ACAGTTCTAGTATTAACATGAACATTTAGAG 2245
QY 2264 AGGAACCTGTGAAATTTAGTGAAGAGTGAAGAACAGTGCMAAGCATTAACCAAG 2323
DB 2246 ATAAACTTTGGCAAAAATAGTTCCGCTGGAAGAAATGCAAGATTAATCAAG 2305
QY 2324 AATTAGACAACGGCGAGCTGATCTAATAAAAAAGTTAAGACATATGAGGAATTAAGA 2383
DB 2306 AACTGTAAGAAAGCAAAATAATGATTTAGAGAAAACTTAAGATTAAGAAAAATTA 2365
QY 2384 AACGTGACAGAGAACATGAACAAGTCACTGTTTGTCTCACTATTAAGAAAAAG 2443
DB 2366 AGGATACAAAAATGCAATGGAAGAAACAAATCTGTTTTTCAACAATTAATCAAG 2425
QY 2444 AAGTATATGATCAAAAAGTAAATAGCAAAAAAGATTAAGATGCGCTTCAACGAGC 2503
DB 2426 ATATATAACGAAAAAG-----AGTCMAAGCTTGAACGCTTGTAG 2470
QY 2504 TTCAAGATACCAAAACATGTGAATAATCTAGGAGGAGTTAAGAGTATCCGTA 2563
DB 2471 TTCAAGATATTTGATCAATTT-----AACTTTGAAAAAGAAATGAAAAAGTTCAAGTGA 2524
QY 2564 CAGAAATTGAAGATTAAGCATTTGATTTGGCAGCAGAAATTTGAAGATATAGATT 2623
DB 2525 CAGAAAAAGAAAGCAAAAGCGTTGATTTGATGACATAGCAATTCAGCTTTATAGAGT 2584
QY 2624 TGAAGAAAGATGTAATTAATTTGAATCAGATTCAGAAATTAAGAGATTGCAAGACT 2683
DB 2585 TAAAAAGAAAGCTGACCATCTAAAGATGATTCGAATTTAGAAAAAGAAATGTAA--T 2641
QY 2684 TAGAAGAAATGTAAGAAAAAGTTAATAGGCTTGCAGATCTGAAGCTCTGAGGTGA 2743
DB 2642 GTAAAGACCAAGTGAAGATTAAGAAAAATATTTTAAATAATGAACCACTGAAGTAA 2701
QY 2744 AGCCGACAGAAACAGTGCAGAAAGTACAAAGCAACTAACAAACAACCAACCGTTG 2803
DB 2702 AGCCACATGAAATTAAGAAAGCTTAACGAAACCAATTAACAAGTGCAGAAACAGTCA 2761
QY 2804 CCGATCCGAGGCAACG-----AATGCAATCTTACAGAACAGACA 2848
DB 2762 AAGAAAGCAAAAAACAGTAGAGACGGAAGAAATGCAAAATCTCTGACCAACAGCA 2821
QY 2849 CATGGGTTACACACATCGACACACACAAAGCAGTCAATCAATCTTACCATCAAT 2908
DB 2822 CGTGGGTCAAAAGCTCAACCAATACAGACCTTCAAGCATGCTTCAAGTTAGGT 2881
QY 2909 CAAAAATTAACATTTGACATCAACGAGCGCATCAACCAACCAAGTGTATGACAGAGAGG 2968
DB 2882 CAABAATTAACCTGACCTCGACGAGCGGTGTAAAGCTTAAGAAATGTATGCAAGAGAGG 2941
QY 2969 AAGATGATGACGAGAGCTGAACCCGAGTGGGGCTGAGATGATGAGGTGGAATGTGA 3028
DB 2942 AAGATGAAGCAGAGAGGTGAAGCCGAGTGAAGGGGCTGAGATGATGAGGTGAGTGTGA 3001

QY 3029 TGAGGGGGGTAGTAGAGCAATGATTTTCTGTCATGATT 3069
DB 3002 TGAGAGGGGTATTATTAAGCAATGATGATTTCAATCATGATT 3042

RESULT 7
US-09-762-724-3
; Sequence 3, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OR INVENTION: Identification of a region of the major surface
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/09/762,724
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3006
; TYPE: DNA
; ORGANISM: *Pneumocystis carinii* sp. f. *hominis*
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3006)
; US-09-762-724-3

Query Match 30.2%; Score 927.4; DB 4; Length 3006;
Best Local Similarity 60.3%; Pred. No. 2,2e-193;
Matches 1859; Conservative 0; Mismatches 1111; Indels 93; Gaps 16;

QY 2 TGCGCGGGCGGTCAACGCGCAGCAGCAGGACACAGAAATAGCATTTGATGAGAACATG 61
DB 2 TGCGCGGGCGGTCAACGCGCAGCAGGCTGCAAG---CACAGAAATAGTGTGAAGAAATATC 58
QY 62 TTTTGAATTTATTTCTAAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 121
DB 59 TTTTGGCTTTGATTTTGAAGAAATGAATGAAGAAATGAATGAATGAATGAATGAATGA 118
QY 122 AAAAAATATTCGCAAGATTTGCTGAACCAAACTAAATATTAAGCAAGTACACAGAAAC 181
DB 119 AAGGATTTTGTGAAGTTTAAAAAATTAACAAAAAGAACAAAACTAGAAAGAAAGT 178
QY 182 TTAAGGTTTTTTCGAAGATGAAAAAGC---AGATACAAAATGCAAAAGCTGAAGCCA 238
DB 179 TAGACGAATCTGCAAGATGATTAATAACAATGAAGCAAAATGCAAAAGATCAAGAACAA 238
QY 239 ATATTTGAAAAAATGATCTACATCAAAAGAAAACTTAAAGAAAGCAATTTAAAAA 298
DB 239 AGGTTAAAGCAAGTGTACTAGTTTCAAAACGAATCTGTAAAGCAGTCAAAAAAGGAG 298
QY 299 TTCAAGTTATTAAGCATTAAGATTTGCAAGGAATGAACAACAATGCTATTTTGAAG 358
DB 299 CTTCAACATTAAGAAATTAAGATTTGTAAGAAATGAACAACAATGCTATTTTGAAG 358
QY 359 GAGTATGTTCAAAAGAACTTAAGATTTGCAATTAATTTTGAAG 418
DB 359 GAGCATGTCACAAAGAACTTAAGATTAATTAAGTAAATGAAGTGAAGTAAATTTATCAAA 418
QY 419 AGAAAGCTGAATTAAGTTCGGAAGAAAGTCTTTTAAAGCACTTCGTAAGCATTTTAA 478
DB 419 AAAAAAGAGAGAGTGAAGCAAAAAAGCTTTTAAAGTAAAGTAAAGTAAAGTAAAG 478
QY 479 GATCAGTCATATGTAAGAAAAAATTAAGAGATTTGCTGTCATGAGGAGGAAAGT 538
DB 479 ATAAAAACATGCAAAATTAAGTAAAGGAGTTTGTCAAGAAATTAACAAGAAAGT 538
QY 539 ATGAGTTAACAACTTGTGTGTAACCAAGAAAGACATGTAAGAAATTTTAAATGA 598

539 ATGAGCTAATAAATTATGTCTTGACGAAGAAAAACGTGTGAGATCTTGTATCTAAGA 598
OY 599 AAGATAGAAAGTCGCGTACTCTTAAACAGATGTTTCAGCAGCACTAGAA-----GTT 652
DB 599 AAGAAATACAAATGCAAACTCTCAAAAGAAAGAAATGATCTAGTGTCTTGAAAGAAAGATT 658
OY 653 TTTAAAAAAGAAACATGTCTTGAATTATCTCGAAACAAATGCTATTTTAACTTGGAAATTGCG 712
DB 659 TATTAAAAAGAAAAAGTTTATTTATTTCTTGAAAGAAATGTTACTTTTATGCGTCAACCTGTG 718
OY 713 GAGACGACGATTAATAATGATGATTTGAAATGGAGAAAAATGCCAAGAACAAATATTG 772
DB 719 AAACGATCAGCA---AAGTATAAGAGTTTGACAGCAAAATGTCAAAAGAAAATCTCG 775
OY 773 CTATATGCCACCAAGACCCGATTTTGATCCAACTAGGCCAGAGCTCAATAGCAGAG 832
DB 776 TTTATGACGACACAGGTTCACTTTGATCTCAACAAATTAAGATTAGTTAGTCAGAG 835
OY 833 ATATAGGCTGGAAGATTTTATAGAAAGTGAAGAGATGAGAGATTTTATTTGAAAGA 892
DB 836 AAATGACCTAGAAAAAATTTGATAGTAAGACAGTAAAAAGGAATTCATATTGGAAGC 895
OY 893 ATCATCTAAGAGATCGACAGCTTTGTTGGCATTTGGATCCAAAGATTCTAGTCTTAA 952
DB 896 CATCAATTAAGATGAAGTCGCTTTATTTGGCATTTTAAGCA-----GAGTGAATGCTC 949
OY 953 AAAAAAGCAGCAAGAAAGATGCGAAGAGCCCTTCAAAAAAGCTGCAAAAATCTCTATG 1012
DB 950 AAAATACCTTTTAAAGATCAATGTAAGATGTTATTTAAAAAAGTGGAAAACCTTTAAG 1009
OY 1013 AACATGAGGCTTTAAGATTTATGTAAGAAAAATGTTAAGTAATGATGGAAGAAAA 1072
DB 1010 AGCATATTTATTTTAAAGATTTATGTAAGTAATGATCACTATCATATTCAAAAAGAA 1069
OY 1073 AATGGAAGAAATGCAAAATGATATTAACAAAATTCGAAAAATTTCACTTCAAAAGTCA 1132
DB 1070 AATGGAAGAACTAATTAAGAGGT-----AACAAACCGTATTTTAACTGTTCTA 1123
OY 1133 CTAAATATGCTCTTTTGAATCCAAACAAAGAAATTAAGAAATTTGTTGATGGAGGCT 1192
DB 1124 GGATTTGAAGAAATTTCCCTCCAGCTAATGTAAA---GGAAATTTATGTTGGCATATGT 1180
OY 1193 TGCCAACTTTCTTTCGAAACGAAGATGTGCGAAATTTGAGATCCATTTGTTCTATTTG 1252
DB 1181 TGCATACATTTCTTGTGTAAGAGAGATGACGAACTGTTGCGAATGTTTAAATTTGA 1240
OY 1253 AAAAAAAATGTCAGATGAGAAATGATGTAATAATTAAGACCAATGTTCAAAA 1312
DB 1241 AAAGCAAGCTCACTT---GAAAAGCCCTGCAATTACTTAAACAGAGATGTTTAA 1297
OY 1313 GAGGATTTGATGACGCGCAATTAAGTGTGCAAGAAATTAAGCAAGATGTTACATG 1372
DB 1298 AAGGCTTTGAAGCAATGAAATGAAGCATTAAGATTAAGGCGGAAAAATTTGCAAG 1357
OY 1373 GTTCAACAAAAGCTGGCTGAAAAGTTTCAACAAGATTTGTAAGAGATGTAAGAA 1432
DB 1358 GTTCAATTAAGCAATGAGCTGTAACCTTCAAAAAAATGTTAAAGTTGTGAAAA 1417
OY 1433 TGAAGAAAAAACAAGAAAGTTTCTCAACAGATTAATTTATTTCTGTGTGTAAGC 1492
DB 1418 CGAAAGAGAG-----AAGTGAAGATTAATTTGTATCTATGTAAGAAC 1459
OY 1493 CAGCAAAAGACCCGCTTCTTCAACATGATCTTGAATGAAACATATCTTTTACAGAC 1552
DB 1460 CAATTAATAACGCGCTTACAGTGTCAACAGATTTGCAATGAGGCGAGTGTCTTGCAAG 1519
OY 1553 AACAACTGATCAAAAGCAGATTTCCCGACAGATTAATAATGCAAGAAATTTGGGAGAA 1612
DB 1520 AGCATTTTGAAGAAAAAGAGATTTTCAACAGAAAAAGATTTGAAGATTAAGAGAAA 1579
OY 1613 AGTGCAGAAATTTAGAGAGATTCAAAAAGAAATTAACATGCGCATGCTACATCGAGAC 1672
DB 1580 AATGTGAGGCTTTAGAGAAAAAGATTCAGAGAAATTAATGTGTCAATGTTATAGCTTAAAC 1639

OY 1673 AGCATGCAATCGCTTGGGACCTACAGAAATTTTAAAGCAGTTTATTTGATGACACA 1732
DB 1640 AGCATTCGAATCGGCTCAGAGAGCATAGACCTTAGAAGAGAGATTTGTTAAAGAAATA 1699
OY 1733 AAGATCTTTGAAAGACCAAGAAAGTTGTGTAATATCTTAAAGAAAGTGTATAAAT 1792
DB 1700 AAGATATTTTAAAGATGAAGAAATGCTGCAAGAAAGAGCTTAAGAAAGATGTGAAAAT 1759
OY 1793 GGTCTAGAGAGAGATGACCGTTCTCTTTGTATGTCTTCAAAACGCTACGTGTG 1852
DB 1760 GGTTAGAAGAGAAATTAATTAATTTTTCGCTGTGCTGACTGGAACCTTGTTGCA 1819
OY 1853 AGCTGATGTTAAAGACGTGAAGACAGGTGTGAATTTCAAAAAAATTTAAAGCTT 1912
DB 1820 AAAAGATCACTAGAAATGTGAATCTAATGTATTAATTAAGAAAGCATATGGAACCTA 1879
OY 1913 CATATATTAATGAATTTCTTGAATATATCAATTAATAACAACCTGAAAGAAAT 1972
DB 1880 TGAACTTTATAGTGAATATGCTTAAAAAGAGAAAAATATGTGAAT-----TT 1930
OY 1973 GTCCCTTGGCATACGATATTCGAATGATTTTCACTAATTTGTCAGGTCTTACGAAG 2032
DB 1931 GGGCTCCATATTTTAAAGATGACAGCAAAATTTGAAAAAATTTAAAAACGAGAGAAAG 1990
OY 2033 AGAATGTTGTACAAAATCAAGAGCATGTGAGCGCTTCTATTAAGAAAGGCTTTGG 2092
DB 1991 ATGGGCAATGCAAAAACTCAATTAAGAAAGTCAATCAATCTTGAAGAAAGAGCTTTAG 2050
OY 2093 AAGATGCTCTCAAGATGAGGCTTCAGGAAATTTGACATGAATATCTTAATGTGAACTG 2152
DB 2051 AAAATTAAGTTGTGAAGAAATTTGAAGAGTATTAATCAACGTAGAGAAATGTAAATTA 2110
OY 2153 CATTGAAGATTAATTTGACATGACGCGGAAACGTAAATATGCTCAATCAGTGTAT 2212
DB 2111 CACTTAATATATCTATGATACAAATTTGAAGAAAGCAGAGATGGGTGGAACCTTGT--- 2167
OY 2213 GCAAGCTACACCAAGATTAATCTTGAAAGATGATGAGATGCTTAAGAAAGAACTCT 2272
DB 2168 -----GCAAAAGCAAGAAACAC---CAAGATGACATTTAAAGTTAGAAAGAACTCT 2218
OY 2273 GTGAAGATTAATGTAAGAAAGATGGAAGACAGTCAAGACATTTACAAACAGAAATTAAGAC 2332
DB 2219 GTGAAGATTAATTAAGCTATTAAGAAAGAAATGCTCAAAATTTGAAGACAGCTTGAAG 2278
OY 2333 AACCGCAGCTGATCTAATAAAAGATTAATTAAGATATGAGAACTTAAGAAACGTGACG 2392
DB 2279 AAGTAAAAAGAGCTTTAGAAAGAAAGAAAGAAAGTATTAATAAAATTAAGAAAGACGAG 2338
OY 2393 AGGAAGCAATGAACAGTCCAGTCTTGTGTTGTCACATTAAGAAAAACGAAGTATG 2452
DB 2339 AAAAAAGCATGAAGATGCAACCTTATTTTATGAGAGCGAAAGGACCTGATTAATAT- 2397
OY 2453 TATCAAAAAGTATATGCAAAAACAAGATTAAGATGCGGTTTCAAAACGGACTTCAAGATA 2512
DB 2398 -----AATTAATTAAGTCAATTAATTAAGACATGATTAACCTTAAGAGAAAGGCA 2452
OY 2513 CCACAAAACATGTGAATAATTAAGAGAGGAGTTAAGATGATCCGTAAACAGATTAAG 2572
DB 2453 CAACAGAGATT-----AACTTGTAAAGAAATGCAAAAGTGACATTAACAGAAAAAG 2506
OY 2573 AAGCTAAAGCATTTGATTTGGCAGCAAGATTAATTTGAAAGATTAAGATTTTGAAGAAA 2632
DB 2507 AATTAAGCAGATTTGATTTGTATCAGAGAGATTAATTTGATCTTAATTAATTTGAAGAAA 2566
OY 2633 GATGTAATTAATTTGGAATCAATGATGCAAGATTAAGAGAGATTTGCAAAAGCTTAAGAAAG 2692
DB 2567 TATGTAATCAATCTACGAAGATTTGTGTTTCAAAAAAGATG---TACTGTGAGGATC 2623
OY 2693 TATCAAAAAGATTAATTAAGCTTTGTGCAATCTGAAGCTCTGAGAGTGAAGCGCAGC 2752
DB 2624 CATGTAATAAAAGATTAAGAGGATTAATGTTCAATTAAGGCACTAATAAGTGAAGACACAGC 2683

QY	2753	AAAGAGGACGAAAGTACAAACGACCACTACAAACA-----ACAAACAAGCTTCCG	2806
Db	2684	AAATATGTAACCTAAATAACATACTACACACAAACACACACCACTACCACTTTAAAG	2743
QY	2807	ATCCGAAGGCAACCGAATGCAATTCCTTACAGACAAACGACACATGGGTTACACAGACAT	2866
Db	2744	ACGCAAAAGGCACACGACTGCCACTCTTTACAGACAAACGATACGTGGGTCAAAAGCGT	2803
QY	2867	CGACACACACAAAGCAGCTCTACTATCAATCTTACATCATCATCTCAAAATAATCAATTGACAT	2926
Db	2804	CGAACCCATATACGACATCCACAAACACATCTTACATGTCAGTCAAGATTAACGTTGACCT	2863
QY	2927	CAACGAGCGCATGCAAAACCAACCAAGTGTACGACAGAGAGAAAGATGATGACAGAGACG	2986
Db	2864	CGACAAACCGGTGTAAACCTACGAAGGTATACGACAGAGAGAGAAATGAAACAGAGAACG	2923
QY	2987	TGAAACCGAGTGAAGGGGCTGAGGATGAGATGGGTGGAAATGTGATTAAGGGGGGTATAGTAG	3046
Db	2924	TGAAACCGAGTGAAGGGGTGAGGATGAGATGATGAGATGATGAGATGAGGGGGGTGTTATTAG	2983
QY	3047	CAATGTTATTTGCTTCATGATT	3069
Db	2984	CAATGACATTTCAATTCATGATT	3006

RESULT 8

```

US-07-781-034-1
Sequence 1, Application US/07781034
Patent No. 5442050
GENERAL INFORMATION:
APPLICANT: Fishman, Jay A.
TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
TITLE OF INVENTION: Rat- and Human-Derived Pneumocystis Carini
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/781,034
FILING DATE: 19911018
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MGH91-02A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2814 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1617..2813
US-07-781-034-1
Query Match 8.5%; Score 261.6; DB 1; Length 2814;
Best Local Similarity 51.3%; Pred. No. 1.3e-47;

```

	Matches	87%	Conservative	0	Mismatches	76%	Indels	6%	Gaps	9%
QY	25	GCACGAGGACACAGAAATAGCAATTTGATGAGCAACATGTTTACGTTTAATTCTTAAAGAA	84							
Db	1152	GTCACAGTACACAGATGAGATTTAAGAGAGAAACACTTTTGCTTCAATGGAGAAC	1211							
QY	85	GATGACTAAGTAGCGCAATGCCAAAAAACTAAAAAATATGGCCAAAGATTGACT	144							
Db	1212	AAACATGATGATGAGATGATGATCAAAAAAGGCTCGAGGAATATGTAAAGCTTGAG	1271							
QY	145	GAAGCAAAACTAAATATAGAAACAATGACACAGAAAACCTTAAAGGTTTTGCGAAGATGA	204							
Db	1272	AAAGCAGATGAGATTT---CAGTGTGATAGAGAAAGTTAAAGACCTTTGTGATGATAA	1328							
QY	205	AAAGCAGATACAAAAATGCCAAAGAACTGAAAACCAATATTGAGAAAAATGTACTACATC	264							
Db	1329	AAACGAGACGAAAAATGTCAAAGAACTGAAAAAAAAGTTAAAGATGAAATGGGAACTTTT	1388							
QY	265	AAAGAAAACTTTAAAGAGCAATTTAAAAAAAATTCAGTTATTAACGGATTAAGATTGC	324							
Db	1389	GATACGATTTGAAAGCATGGT-----AGATGACTATGAAAGATGAGAGATTGT	1439							
QY	325	AAAGAGATAGACAAACAATGCTATTTTTGGAGGAGTATGTTCAAAAGAACTTAAAGAT	384							
Db	1440	AAAAACATAGAGAAAAATGTATCTTTAGAGAGACAGCCCAATAGCTTTAAGAG	1499							
QY	385	GATTGCATATCTTTGAGAAATTAAGTGTATCAAAAGAAACGTATTAAGTTGCGAGAA	444							
Db	1500	AACGTGTCAAGTTTGAAGGAGAGATGTATCGAATTAAGGCTAAAAAGTGGCAGAGAG	1559							
QY	445	GTTCTTTTAAAGACACTTGTAGCGATCTTAATGATCACTCATGTGAAAAAACTT	504							
Db	1560	CTCCTTTTGGGGCGCTCGAGGGGATGCTAAAGATGAACTAAATCTAAAGAAAGATG	1619							
QY	505	AAAGAGATTTGGCCCTGTCATGGGGAGGAAAGATGATGTTAACAACTGTGCTGAC	564							
Db	1620	AAAACTGTTTGCCCAATGTTTAAAGCCGAGAAAGTACGAGCTGATGTTTCTGCTTGAT	1679							
QY	565	CAGAAAGACATGTAAGATATTTTAAATGAAAAAGATTAAGATGCGGATCTTTAA	624							
Db	1680	TGCGATGGAACGTGTAAAGCGCTGAAAACAAATCAGAAAGATTTGCGCTTTAA	1739							
QY	625	ACAGATTTTACAGACACTAGGAAGTTTAAAAAAGAAACATGCTTGAATTACTCGAA	684							
Db	1740	GAAAAAGCTTAAAGTGGCGAAT-----AAAGAAAAATGTCATGAAAGCTTGAG	1790							
QY	685	CAATGCTATTTTAACTATGGAATTTGGGAGACGACATATTAATTAATGATTTGAATG	744							
Db	1791	AAATGCTATTTTAAACAAAGAGCGGTATGA-----ACAAAGTGTGATGAGAT	1841							
QY	745	GAGAGAAATATGCCAAGAACAAATATTTGCTTATATGSCACAGAACCGAATTTGATCCA	804							
Db	1842	ATGAGAGATATGCAAGAAAAAGATTCACATATTAAGCGCGGAATTTGATTTAGTCTT	1901							
QY	805	ACTAGGCCAGAGGCTACAAATAGCAGAGATATNAGGCTGGAAGATTTTAAAGAGTA	864							
Db	1902	GTCAGCCGAAAGCGTGTGTTGAGAAAGTATGGGTGATGATGTATTAATAAAGGCT	1961							
QY	865	GAGAGGATGAGATTTTATTTGGAAGATCATCTTAAGATGCGACAGCTTTGTGGCA	924							
Db	1962	GAAAAAGAAAGAAATTAATTTAGAAAATCAGAGGTGATCTAACAGAACTCAGTTACA	2021							
QY	925	TTGTGATTCAGATTCCTAGTCTTAAAAAAAAAAGACAAAGAGAAATCGAAGATGCC	984							
Db	2022	AAATTTCTGCAAGATCTCTTGCTACTGTTGACAGAGATGAGATGTCAGGAGAGAA	2081							
QY	985	CTTCAAAAAAGCTGCAAAAAATCTCTATGAAACATGAGGCTTTAGAAAGTTTATGTAAGAA	1044							
Db	2082	TGGGTATTAACGTTTAAAGAAAATGTGAACCTTATGATTTGAATCTGATTTGATGAG	2141							
QY	1045	AATGCTTTAAATATGATGAGACGAAAAAATGTGAAGATTTGCATAATGATTAACAA	1104							
Db	2142	TTATGCAAAAGATCTGTATTAAGAAAAATTAATGCAAAAAAAGCTTAGATG---TAAAGAA	2198							

QY 1105 ACTTGCAAAATTTTCATCTTCAAAAGTCATTAATTCCTTTTTCATCCAAACAAAGCA 1164
 Db 2199 AGATGTACAAAACCTCAAGTTAAATCTTATGTGAAGGGTGTCTACCGAGTTTAAAGAA 2258
 QY 1165 AATAATGAA-----ATTGTTGATGGGAAGGGTGGCAACATTTCTTAGCAACGAAT 1218
 Db 2259 GATTAATAATCATCATCTTTTATCGTGGGACAGCTTCCAACTTTATTTACGAAGGAGAG 2318
 QY 1219 TGTGCGAAATTTGAGTCTTATGTTTCTTATTTTGAATAAAATGTCAGATGAGAA-- 1276
 Db 2319 TGTGAGAACTTGAATCGGAATGTTTCTATTTAGAAATGCGTGAATAGATAGAGTT 2378
 QY 1277 ----ATGCATGTAAATATTAAGCAACATGTTACAAAAGAGAGCTTGATGACGGCA 1332
 Db 2379 GGTGAAGCGTCAAAATCTACGATCAGCTGCTATTAATAAGGAGCAAGACAGATGTTG 2438
 QY 1333 AATAAGTCGCAAGAAATATGAGGAAGT--ACATGTTCAACAAAAGCTGG 1389
 Db 2439 AATAAGTCTTTCAAAGGAATTAAGGAAAGCTTGTGATGAATTTTATAGCGAT 2498
 QY 1390 CTGGAAGATTTCAACAAGAAATTAATAAAGTATGTGAGAACTGAAATAAAAGAAACAA 1449
 Db 2499 CCTAAGATTTGTAAATAATATGTGTAGAAACGTACAAAACCTTAAATAAGATTAAGA 2558
 QY 1450 GGAAGTTCTCAACAGATGAATTTATTTCTGTGTGACAGCCAGCAAAAGCAGCCCG 1509
 Db 2559 TACCTTCAAAA-----TGTCTTATCTTAAAGAACTATGTTAT 2597
 QY 1510 TTGCTTACACATGATCTTGAATGAATACTATCTTTTACGACAACTGGATCAAAAAG 1569
 Db 2598 GGGCTTCAAAAGATTAATTTTCTCCAAATCCAAAGATTAGTCCCTTTTATGATATCAG 2657
 QY 1570 CGAGATTTCCCGACAGATTAATAATTTGCAAGAAATTTGGGAGAAAGTCCCAAGATTAGCA 1629
 Db 2658 AGAGATTTTCCATTTGAAAGAGATTGTCTGAATTTGGGAGAAATGTGATCAACTTAGT 2717
 QY 1630 GAGATTTCAAAAGAAATTAACATGCGCATGTCTATACCTGAGAGCAAAATGCAATGCTTG 1689
 Db 2718 AGTATTTCAATTAATTAAGAAATTTAGAAAGTGTATTAACATGGAAGAGCGCTGTGAATATTTT 2777
 QY 1690 GGGACTACAGAAATTTTAAAGCAGGTTTAT 1721
 Db 2778 GACGTTACGAAAGATTTTGAAGAAATGATTTT 2809

RESULT 9

PCT-US92-08328-1
 ; Sequence 1, Application PC/TUS9208328

; GENERAL INFORMATION:
 ; APPLICANT: Fishman, Jay A.
 ; TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
 ; TITLE OF INVENTION: Rat- and Human-Derived Pneumocystis Carinii
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 ; STREET: Two Militia Drive
 ; CITY: Lexington
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02173
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US92/08328
 ; FILING DATE: 19920930
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/781,034
 ; FILING DATE: 18-OCT-1991

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/768,166
 ; FILING DATE: 30-SEP-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Granahan, Patricia
 ; REGISTRATION NUMBER: 32,227
 ; REFERENCE/DOCKET NUMBER: MGH91-02AA PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 861-6240
 ; TELEFAX: (617) 861-9540
 ; TELEX: 951794
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2814 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1617..2813
 ; PCT-US92-08328-1

Query Match 8.5%; Score 261.6; DB 5; Length 2814;
 Best Local Similarity 51.3%; Pred. No. 1.3e-47;
 Matches 879; Conservative 0; Mismatches 764; Indels 69; Gaps 9;

QY 25 GCACAGGAGACAGCAATAGCATGATGAGGACATGTTTATGCTTAATCTTAAGAA 84
 Db 1152 GTACAGAGACCAAGATGAGATTAAGAGGACACCTTTGGCTTTCATTTGTAGAGAC 1211
 QY 85 GATGACTTAAGTGAAGCAGATGCAAAAATAAAATAATGTCAGAAATGACT 144
 Db 1212 AAACATGATGATGAATGATGATGCAAAAAGCTCGAGAAATTTGTAAGAGTTGAG 1271
 QY 145 GAAGCAAACTTAATATTAAGCAAGTACAGAAACTTAAGTTTTCGGAAGATGA 204
 Db 1272 AAACAGATGAAATTT---CAGTGTGAATGAGAAAGTTAAAGACCTTTGTGATGATGATA 1328
 QY 205 AAACAGATCAAAATGCAAAAGCTGAAAGCCAAATTTGAGAAATAATGATACATC 264
 Db 1329 AAACAGATGCAAAATGCAAAAGCTGAAATAATTAAGATGAAATGGAACCTTTT 1388
 QY 265 AAAGAAACTTAAGAAAGCAATTAATAAAATAATTCAGATTATTAACGATTAAGATTGC 324
 Db 1389 GATACGATCTTGAAGCATGGT-----AGATGACATGAAGATGAAGAGTTTGT 1439
 QY 325 AAAGAAATGAACAACAATGCTTATTTTGGAGGAGTATGTTCAAAAGAACTTAAGAT 384
 Db 1440 AAAAACAATGAAGAAATGATTAATCTTTTAGAGGAGACGACCCAAATGATTAAGAG 1499
 QY 385 GATTGCAATCTTTGAAGAAATTAAGTGTATCAAAAAGAAAGTGAATTAAGTTGCGGAGAA 444
 Db 1500 AACTGTGCAAGTGAAGGAGAGATGTTGCAATTTGAAGCGTAAAGGTGCGAGAGAG 1559
 QY 445 GTTCTTTTAAGAGACTTGTGAGCATCTTAATGATCATGATATGTAAGTGAATAAACTT 504
 Db 1560 CTCCTTTTGAAGGGGCTGCGAGGGAGTCTTAAGATGAAGTAAATGTAAGAAAGATG 1619
 QY 505 AAAGATTTGCCCTGTGATGAGGAGGAAAGTATGATTAACAACTTGTCTGAAC 564
 Db 1620 AAAACTGTTGCCCAATGTTAAGCCGAGAAAGTACGAGCTGATGTTTTCGCTTGAT 1679
 QY 565 CAGAAAGACATGTAAGAAATTTTAATGAATAAGTGAAGTGGCGATGCTTTAA 624
 Db 1680 TCGATGGAACGTGAAGCGCTGAACAAATCAAGAAAGTTTGCTGCTTAATAA 1739
 QY 625 ACAGATGTTTCAGACGACTAGGAAGTTTAAATAAGAAACATGCTTGAATTAACGAA 684
 Db 1740 GAAAGCTTAAGATGGCAATTT-----AAAGAAATATGATGAAGAAAGCTTGAG 1790

QY 685 CAATGCTATTTTACATGGAATTCGCGAGACGACATATTAATTAATGATTAATG 744
 DB 1791 AAATGTATTTTTCACAAAGAGCGTGACTGA-----ACAAAGTGTGATGAGAT 1841
 QY 745 GAGAGAAATTCGCAAGAACAAATATGCTTATATGCCACGAGACCCGATTTTGATCCA 804
 DB 1842 ATGAAAGCAATGCAAGAAAAGATTCACATATTAAGCGCGAATCTGATTTAGTCTT 1901
 QY 805 ACTAGGCGACAGGCTACATGACGAGAGATATAGGCTGCGAAGATTTATTAAGAAGTA 864
 DB 1902 GTCAAGCGAAGGGGTGGTGTGTGAGAAAGTATGGGTGGATGTATTAAGAAAGCTT 1961
 QY 865 GAGAGAGATGAGATTTTATTAATGAAAGATCATTAAGAGATGCGACAGCTTTGTGGCA 924
 DB 1962 GAAAAAGAGAAATTTATTTTGGAAAATCAGAGATGATCTACCAAGAACTCAGGTACA 2021
 QY 925 TTGTTGATCCAGATTTCTAGTCTTAAAAAAAAGACGACAAAGAAATGCGAAGAACCC 984
 DB 2022 AAATTTCTGCAAGATCTCTGCTACTGTGTGAGCAGAGATGAGATGATGCGAGAAAGAA 2081
 QY 985 CTTCAAAAAGCTGCCAAAAATCCCTCATGACATGAGCTTTAGAAAGTTATATGAAGAA 1044
 DB 2082 TGGCGTAAAGCGTGAAGAAATGAAACTTCTAAGATTTTGAATCTGATTTGATGAG 2141
 QY 1045 AATGTTTAAGTAAATGATGAAACGAAAAATGTGAAGATTCGAAATGATATTAACAAA 1104
 DB 2142 TTATGCAAAAGTGTGATTAAGAAATTAATGCAAAAAAAGCTAAGT---TAAAGAA 2198
 QY 1105 ACTTGCAAAATTTTCACTTCAAAAGTCACTAATAATGCTTTTGTATCCAAACAAAAGA 1164
 DB 2199 AGATGTACAAAATCAAGTAAATCTTATATGAAAGGGTGTCTACGAGATTTAAAGAA 2258
 QY 1165 AATTAATGA-----ATTGTGATGAGGAGAGGTTGCCAATTTCTTGAACAAGAAAT 1218
 DB 2259 GATTAATAATCAACATCTTTATTCGTGGGACAGCTTCCAAATTAATTAACAAAGGAGAG 2318
 QY 1219 TGTGCAAAATGAGTCTATTTGTTTCTAATTTTGAAAAAAATGTCAGATGAGAGAA-- 1276
 DB 2319 TGTGCAAAATGAGTCTATTTGTTTCTAATTTTGAAAAAAATGTCAGATGAGAGAA-- 1276
 QY 1277 ---ATGCAATGAAAAATTAAGCAACATGTTACAAAGAGAGCTTGTATGACGAGCA 1332
 DB 2379 GGTGACGCTGTCAAAATCTACGATCAAGCGTATTAAGAGAGAGAGAGAGAGATGTTG 2438
 QY 1333 AATTAAGTCTGCAAGAAATATGCGAGAGATGTT--ACATGTTCAACAAAGCTGG 1389
 DB 2439 AATTAAGTCTTCAAAAGAAATTAAGGAGAGAGCTTGTCTATGATTAAGATTTTATAGGAT 2498
 QY 1390 CTTGAAAGTTTCAACAAAGATTAAGATATGATGAGAACTGAAGAAAGAAACAAA 1449
 DB 2499 CTTAAGATTTGTAAGAAATATGATGAGAACTGTAACAAACTTAAAGAAAGATTAAGAA 2558
 QY 1450 GGAAGTTTCTCAACGATGATTAATTTATCTGTGTGTAACGACGACGAAAGACCCCG 1509
 DB 2559 TACCTTTCAAAA-----TGTCTTATCCAAAGATTAATGTTAT 2597
 QY 1510 TTGCTTACACATGATCTTCAATGAAAATCTATCTTTTGAACAACAACCTGATCAAAAG 1569
 DB 2598 GGGCTTTCAAAATGATTTTCTCAATCCAAAGATTAAGTGTGCTTTTGAATGATCAG 2657
 QY 1570 CGAGATTTCCGCAAGATTAATTAAGAAATTTGGGAGAAAGTGCGCAAGATTTTGA 1629
 DB 2658 AGAATTTTCAATTTGAAGAAAGATTTGTTGAAATTTGGAGAGAGATGATCAACTTATG 2717
 QY 1630 GAGATTTCAAAAGAAATTAACATGCGCATATCACTGAGACGACATGCAATGCGCTTG 1689
 DB 2718 AGTATTTCAATTTGAATTAAGAAAGTATTAACATTAAGAAAGCGCTGTAATATTTT 2777
 QY 1690 GGGATCAAGAAATTTTAAAGCAGTTTAT 1721
 DB 2778 GACGTTACAGAAAGATTTTAAAGATATTTT 2809

RESULT 10
 US-07-781-034-3
 : Sequence 3, Application US/07781034
 : Patent No. 542050
 : GENERAL INFORMATION:
 : APPLICANT: Fishman, Jay A.
 : TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
 : NUMBER OF INVENTION: Rat- and Human-Derived Pneumocystis Carinii
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 : STREET: Two Millitia Drive
 : CITY: Lexington
 : STATE: Massachusetts
 : COUNTRY: USA
 : ZIP: 02173
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/07/781,034
 : FILING DATE: 19911018
 : CLASSIFICATION: 424
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Granahan, Patricia
 : REGISTRATION NUMBER: 32,227
 : REFERENCE/DOCKET NUMBER: MGH91-02A
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (617) 861-6240
 : TELEFAX: (617) 861-9540
 : INFORMATION FOR SEQ ID NO: 3:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1448 base pairs
 : TYPE: NUCLEIC ACID
 : STRANDEDNESS: double
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : HYPOTHEICAL: NO
 : ANTI-SENSE: NO
 : US-07-781-034-3
 Query Match 6.28; Score 192; DB 1; Length 1448;
 Best Local Similarity 54.1%; Pred. No. 1.9e-32;
 Matches 565; Conservative 0; Mismatches 440; Indels 39; Gaps 7;
 QY 35 CACAGATAGCATGATGAGGAAATGTTTATGCTTAAATCTTAAGAAAGATGACTAA 94
 DB 432 CACAGATGATTAAGAGAGAACACTTTTGCTTCAATTTGAGAGACAAACATGATG 491
 QY 95 GTGACGAGAAATGCAAAAAAACTAAAAAATTTGCCAAGATTAAGTGAAGCAAAAC 154
 DB 492 ATGAGATGATGATGCAAAAAAGCTCGAGAAATTTGTAAGAGTTTGAAGAAACAGATG 551
 QY 155 TAAATATGAACAAGTACAGAAAACTTAAGGTTTTTGGAGATGAGAAACAGATA 214
 DB 552 AGAATTT---CAGTGAATGAGAAAGTTAAAGACCTTGTGATGATTAAGAAACAGAGCG 608
 QY 215 CAAATGCAAAAGACTGAAGCCATATTTGAGAAAAATTTACTATCAATCAAGAAAGAAC 274
 DB 609 AAAAATGCAAAAGACTGAAGAAAAAAGTTGGGATGAAATTTGGAACTTTTGATCGGATC 668
 QY 275 TTAAGAAGCAATTAAGAAAAAATTTCAATTAATTAAGATTAAGATTTGCAAGAGAAATG 334
 DB 669 TTGAAGATCGGT-----AGATGACATGAAGATGAAGAGATTTGTAAGAAACAGATG 719
 QY 335 AACCAAGATGCTATTTTGGAGGAGATGTTCAAAAGAACTTAAGATGATTTGCAATA 394
 DB 720 AAGAAAAATGATTAATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
 QY 395 CTTGAGAAATTAAGTCTATCAAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 454

```

Db 780 AGTTAGGGAAGATGTTACGAATTTGAAGCCGTAAGGCGTGAAGAGAGCTCTTTTGA 839
Qy 455 GAGCACTTCGTAGCGATCTTAATGATCAGTCATATGTGAAGAAAACTTAAGAGATT 514
Db 840 GGGCGCTCGAAGAGAACTTAAGAAAGTTAAATGTAAAGACAGATGAAGAAAAAGTTT 899
Qy 515 GCGCTGTATGGGAGGAAAGTGTAGTTAACAACTTGTGTGTGAACAGAAAGGA 574
Db 900 GCCCAGTGTAAAGCGAAGAAAGCGAATGATGTTTGTGCTTGAATTCGATGGAA 959
Qy 575 CATGTAAAGATTTTAAATTTGAAGAAAGTAAGAGTGGGATCTCTTAAACAGATGTT 634
Db 960 CGTGTGAG---CGTGAAGAAAAAATCAGAAAGATTTGGCCAGCTTTAAAGAAAGCTTA 1016
Qy 635 CAGCAGACTAGGAATTTTAAAGAAACATGTCTTGAATTAATCTGCAACATGCTATT 694
Db 1017 AAGATG-----GCGAATTTAAAGAAATGTCAATGAAGAACTTGAGAAATGTCAAT 1067
Qy 695 TTTACATTGGAATTTGGGAGACGACGATTAATTAATGTATTTGAATTTGGAGAGAAAT 754
Db 1068 TTTAC-----GGAAGAGCGTGTATTAAGAAATGTGATGAGATTAAGATCAAT 1118
Qy 755 GCCAAGAACAAATATTTGCTTATGTCACAGACCCGATTTGATCCAACTAGCCAG 814
Db 1119 GCGAGAAAAAAGAAATCAGATATAG-CGCCAGAACTGATTTCTGCTGCAAGCCGA 1177
Qy 815 AGGCTACAATAGCAGAGATATAGGCTGTGAAGACTTTTATTAAGAAAGTAGAGAGATG 874
Db 1178 AGAGCTCGTGTGTGAAGATATTTGGTGTGATGATGTGTATTAAGAACTGAAAGAAAG 1237
Qy 875 GAGTTTTTATTTGGAAGATCATCTAAGAGATGCGACAGCTTTGTGGCATTTGTGATCC 934
Db 1238 GAAATTAATTTGGAATCAGAGATGATCTAACAAGAGACAGATCAAAATTTCTGC 1297
Qy 935 AAGATT-----CTAGCTTTAAAGAAAAAGACGAAGAAAGTSCAAGAAAGCCCTTCA 989
Db 1298 AAGATCTCTGTAGTCTTTGAGCAGATGAGATGATAGATCAGAGAAAGAAATGCG 1357
Qy 990 AAAAAAGTGCAGAAATCTCATGAACATGAGGCTTTAGAAAGTTATGTAAAGAAAAATGG 1049
Db 1358 AAAAAAGCTTAAGAAATGTGAAACTTCTAATGATTTGAATGATTTGATGAGATGATG 1417
Qy 1050 TTTAAGTATGATGAGAGAAAA 1073
Db 1418 CAAGATGCTGATTAACAAAAA 1441

```

```

RESULT 11
PCT-US92-08328-3
; Sequence 3, Application PC/TUS9208328
; GENERAL INFORMATION:
; APPLICANT: Flahman, Jay A.
; TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
; TITLE OF INVENTION: Rat- and Human-Derived Pneumocystis Carinii
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P. C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/08328
; FILING DATE: 19920930
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/781,034

```

```

; FILING DATE: 18-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,166
; FILING DATE: 30-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MGH91-02AA PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; TELEX: 951794
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1448 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
PCT-US92-08328-3

Query Match 6.2%; Score 192; DB 5; Length 1448;
Best Local Similarity 54.1%; Pred. No. 1.9e-33;
Matches 565; Conservative 0; Mismatches 440; Indels 39; Gaps 7;

Qy 35 CACGAATAGCATGATGAGGACATGTTTATCTTAATTTCTAAGAAAGATGACTAA 94
Db 432 CACAAGATGATTAAGAGAGAACCTTTTGGCTTTCATTGTGAAGACAAACATGATG 491
Qy 95 GTGACGAGAAATGCAAAAAAATTAATAATTTGCCAAGATTTGACAGCAAAAC 154
Db 492 ATGAGAAATGATGCAAAAAAGCTCGAGAAATTTGTAAAGATTTGAAGAAACAGATG 551
Qy 155 TAAATTAAGAACATGATCAGAGAAACCTTAAGGTTTGGCGAAGTAGAAAGACAGTA 214
Db 552 AGAATTT---CAGTGTGAATGAGAAAGTTTAAAGACCTTTGATGATTAAGAAAGAGCG 608
Qy 215 CAAATGCAAGAACTGAAGCCATATTTGAGAAAAATGTATCAATCAAGAGAAAC 274
Db 609 AAAATGCAAGAACTGAAGAAAAAAGTTGGGATGATTTGGAACTTTGATCGGATC 668
Qy 275 TTTAAGAAAGCAATTAAGAAAAAATTCAGATTATTAAGGATTAAGAGATG 334
Db 669 TTGAAGATCGGT-----AGATGACATGAAGATGAAGAGATTTGTAAAGAAACATG 719
Qy 335 AACAAATATGCTATTTTGGAGGAGATGTTCAAAAGAACTTAAGATGATTCATA 394
Db 720 AAGAAATGTATCTTTTAGAGAGACAGCCAAATAGCTTTAAGAGAACTGTCTCA 779
Qy 395 CTTTGAGAAATAGTGTATCAAAAGAAACGTGATTAAGTTGCGAAGAACTTTTAA 454
Db 780 AGTTAGAGGAAGATGTTAAGAAATTTGAAGAGCTAAGAAAGTGGCAGAGAGCTCTTTTGA 839
Qy 455 GAGCACTTCGTAGCGATCTTAATGATCAGTCATATGTGAAGAAAAAATTTAAGAGATT 514
Db 840 GGGCGCTCGAAGAGAAAGCTTAAGAAAGTAAATGTAAAGACAGATGAAGAAAGTTT 899
Qy 515 GCGCTGTATGGGAGGAAAGTGTAGTTAACAACTTGTGTGTGAACAGAAAGGA 574
Db 900 GCCCAGTGTAAAGCGAAGAAAGCGAATGATGTTTGTGCTTGAATTCGATGGAA 959
Qy 575 CATGTAAAGATTTTAAATTTGAAGAAAGTAAGAGTGGGATCTCTTAAACAGATGTT 634
Db 960 CGTGTGAG---CGTGAAGAAAAAATCAGAAAGATTTGGCCAGCTTTAAAGAAAGCTTA 1016
Qy 635 CAGCAGACTAGGAATTTTAAAGAAACATGTCTTGAATTAATCTGCAACATGCTATT 694
Db 1017 AAGATG-----GCGAATTTAAAGAAATGTCAATGAAGAACTTGAGAAATGTCAAT 1067
Qy 695 TTTACATTGGAATTTGGGAGACGACGATTAATTAATGTATTTGAATTTGGAGAGAAAT 754

```

Db 1068 TTAC-----GGAGAGCGTGTGTATTAACAAATGTGATGAGATTAAGATCAAT 1118
Qy 755 GCCAGAGAAATAATATGCTTATATGACAGAGACCCGATTTGATCCAACTAGGCGAG 814
Db 1119 GCGAGAAAAGAAATACATATTAAG--CGCCAGAACTGATTCATGCTCTGCAAGCCGA 1177
Qy 815 AGGCTACAAATAGCAGAGATATAGGCGCTGAGAAAGTTTATTAAGAGGTAGAGAGATG 874
Db 1178 AGAGCTGCTGTGTAAGAGATTTGGGTTGGATGATGTATTAAGAGCTGAGAAAAGAG 1237
Qy 875 GAGTTTATTTGAGAAAGATCATCTTAAGAGATGCGACGCTTTGGCATTTGATGCC 934
Db 1238 GAATATATATTTGAGAAATACAGAGCTGATCTTACAGAGAGAGTCAAGTAAATTTCTGC 1297
Qy 935 AAGAT-----CTAGCTTAAAGAAAAGACAGAAAGAAAGCGAAGAGCCCTTCA 989
Db 1298 AAGATCTCTTGTCTAGCTTTGCGAGAGATGAGAAATGATTAAGAGATGAGGAGAAATGCG 1357
Qy 990 AAAAGCTGCAAAATCCTCATGAAACATGAGGCTTTAGAAAGTTTATGTAAGAAAAATGG 1049
Db 1358 AAAAGCGTTAAABAAAATGTGAAACTTCTAGTATTTGAATCTGATTTGATGAGTTATG 1417
Qy 1050 TTAAATGATGATGAAACGAAAA 1073
Db 1418 CAAAGATGCTGATTAACAAAAAA 1441

RESULT 12
US-09-762-724-15
; Sequence 15, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/09/762,724
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 249
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(249)
US-09-762-724-15

Query Match 6.0%; Score 185; DB 4; Length 249;
Best Local Similarity 83.9%; Pred. No. 3.6e-31;
Matches 209; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 2821 GAATGCAATCTTACAGACAGACACATGGGTTACAGACATGACACACACAGC 2880
Db 1 GAGTGCACATCTTCAGACAGACACACGCGGTCACAAGAGCTGCACCATATATGAC 60
Qy 2881 ACGCTCTATCAATCATCATCATCAATCAAAATTAATCATGATGATCAAGAGCGATGC 2940
Db 61 ACTTCTAGCATGATGCTCACAGTCACTCATGAAATTAATCACTCAAGAGCGGTGT 120
Qy 2941 AAACCAACCAAGTACAGACAGAGAGAGATGATGACAGAGACGTGAAACCGAGTAG 3000
Db 121 AAGCTTACGAAGTACAGACAGAGAGAGATGATGACAGAGAGGTGAAACCGAGTAG 180
Qy 3001 GGGCTGAGATGAGTGGTGAATGTGATGAGGGGGGTGATAGTGAATGCTTATTTTCG 3060
Db 181 GGGCTGAGATGAGTGGTGAATGTGATGAGGGGGGTGATAGTGAATGATGATTTTCA 240

Qy 3061 TTGATGATT 3069
Db 241 TTCATGATT 249

RESULT 13
US-09-902-540-1357
; Sequence 1357, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1357
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(612)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1357

Query Match 2.9%; Score 87.6; DB 4; Length 612;
Best Local Similarity 46.7%; Pred. No. 1e-09;
Matches 273; Conservative 0; Mismatches 311; Indels 0; Gaps 0;

Qy 83 AAGATGACTAAGTACAGAGAAATGCAAAAAATTAATAATATGCTCAAGATGA 142
Db 4 AA 63
Qy 143 CTGAGCAAACTTAATATGAAACAGTACACAGAAACTTAAAGTTTTCGAAATG 202
Db 64 AA 123
Qy 203 GAAAGCAGATACAAATGCAAGAACTGAAGCCATTTAGAGAAAAATGTACTCAA 262
Db 124 AA 183
Qy 263 TCAAGGAAACTTAAAGAACTTAAAGAAATTTAGATTAACGATTAAGATT 322
Db 184 AA 243
Qy 323 GCAAGCAATGAACAATGCTTATTTTGAAGGAGATGTTCAGAAACTTAAAG 382
Db 244 AA 303
Qy 383 ATGATTCATATCTTGAAGAAATAGTGTATCAAAAGAAACGATTAAGTTGCGAG 442
Db 304 AAAAAAAAAATTAATAATTAATAAAAAAAAAAAAAAAAAATTAATAATAA 363
Qy 443 AAGTCTTTTAAAGACCTTGTAGCATCTTAATGATCAATGATGTAAGAAAAAC 502
Db 364 AA 423
Qy 503 TTAAGAGATTTGCTGTCTGATGAGGAGGAAAGTGTAGCTTAAACAATCTGTCTGA 562
Db 424 AA 483
Qy 563 ACCAGAAAGACATGAAGATTTTAAATGTAAGAAAGTGGCGTACTCTTA 622
Db 484 AA 543
Qy 623 AACGATGTTTCAGCAGCACTAGAAATTTTAAAGAAACA 666

Db 544 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 587

RESULT 14
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300. 6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpc-F18
; US-08-232-463-14

Query Match 2.8%; Score 86.6; DB 1; Length 7218;
Best Local Similarity 2.8%; Pred. No. 3.8e-09;
Matches 11; Conservative 254; Mismatches 128; Indels 0; Gaps 0;

Db 56 AACATGTTTAACTTTAAATCTTAAAGAAATGAGCTAAGAGAGAGAAATGCAAAAAA 115
1456 AAGAGATGAAGAAATTTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1397
116 AACTAAAAAATATGCGCAAGAAATGACTGAAGCAAACTAAATATAGAACAAAGTACACA 175
1396 RRR 1337
176 GAAATCTTAAGAGTTTTCGAAAGTGAAGAAAGATACAAATGCAAGAACTGAAG 235
1336 RRR 1277
236 CCAATATTTGAGAAAAATGCTACTACATCAAGAAAACTTAAGCAAGCAATTAATAAAA 295
1276 RRR 1217
296 AAATTCAGATTATTAACGATTAAGATTGCAAGAGAAATGACAAATGCCATTATTTTGG 355

Db 1216 RRR 1157

356 AGGAGTATGTTCAAGAACTTAAGATGATGCAATCTTGAAGAAATAGTCTATC 415
1156 RRR 1097

416 AAAAGAAACGTGATTAAGTTGCGGAGAGAGTTTC 448
1096 RRR 1064

RESULT 15
US-09-902-540-1280
; Sequence 1280, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(115849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1280
; LENGTH: 1039
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(1039)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1280

Query Match 2.4%; Score 74.4; DB 4; Length 1039;
Best Local Similarity 46.2%; Pred. No. 9.4e-07;
Matches 283; Conservative 0; Mismatches 323; Indels 6; Gaps 1;

Db 70 TTAATTCGAAGATGAGCTAAGTGAAGAGAAATGCAAAAAAATCTAAAAAATAT 129
118 TTAATTAAGAAAAAATATTAAGCAATATTAAGAAAAAATATTAAGAAAAA 177
130 TGCCAGATTTGACTGAAGCAAACTAAATATAGAACAGTACACAGAAAACTTAAGGT 189
178 AATTAAGAAAAAATATTAAGCAAACTAAATATAGAACAGTACACAGAAAACTTAAGGT 237
190 TTTTCGAGATGAGAAAGCAGATACAAATGCAAGAGCTGAAGCAATATTTGAGAAA 249
238 AATATATTAAGAAAAAATATTAAGCAAACTAAATATAGAACAGTACACAGAAAACTTAAGGT 297
250 AATATGTAATTAAGCAAGAACTTAAGCAAACTTAAGCAAACTTAAGCAAACTTAAGGT 309
298 AAAAAAATTAAGAAAAAATTAAGCAAACTTAAGCAAACTTAAGCAAACTTAAGGT 357
310 ACGATTAAGATGCTCAAGAGATGAACAAATGCTATTTTTCGAGAGAGTATGTTCA 369
358 AATTAAGCAAAAAAATTAAGCAAACTTAAGCAAACTTAAGCAAACTTAAGCAAACTTAAGGT 417
370 AAGAACTTAAGATGATTTGCAATCTTGAAGAAATAGTCTATCAAGAAAGCGTAT 429
418 AATTAAGCAAAAAAATTAAGCAAACTTAAGCAAACTTAAGCAAACTTAAGCAAACTTAAGGT 477
430 AATGTTGCGAGAGAGTTCTTTTAAGCACTTGTGAGCACTTTAATAGTCACTATA 489
478 AATTAATTAAGAAAAAATTAAGCAAACTTAAGCAAACTTAAGCAAACTTAAGGT 537
490 -----TGTGAAGAAAAAATTAAGAGATTTGCCCTGTGATGGAGGAGGAGTATGAG 543
538 AAAAAAATTAAGAAAAAATTAAGAAAAAATTAAGAAAAAATTAAGAAAAAATTAAGGTAT 597

QY 544 TTAAACAACTGTGTGTGAACGAGAAAGACATGTAGATATTTTAATTGAAAAGAT 603
| | | | |
Db 598 TTTAAAAACAAAAAANAAAAAANCAGCATAAAAATATAATATTAACATTAATAAAAAAC 657
| | | | |
QY 604 AAGAGTGGCGTACTCTTAAACAGATGTTTCAGCAGCACTAGAGATTTTAAAAAGAA 663
| | | | |
Db 658 AAAAAAAAGTGTTTTCTTAAAAAANAAAAAATCTCCCTAAAAACATAAAAAANAAAAA 717
| | | | |
QY 664 ACATGCTTGA 675
| | | | |
Db 718 AAAAGAAATTA 729

Search completed: October 29, 2005, 09:09:44
Job time : 718 secs

This Page Blank (uspto)

Query Match 100.0%; Score 3072; DB 19; Length 3072;
Beet Local Similarity 100.0%; Pred. No. 0;
Matches 3072; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGGGGGGGTACAGCGGAGGAGGAGACAGAAATGATGATGAGGAGAT 60
DB 1 ATGGGGGGGGGTACAGCGGAGGAGGAGACAGAAATGATGATGAGGAGAT 60
QY 61 GTTTTAGCTTTAATTCTAAGAGAGATGACTAATGAGCAGAAATGCAAAAACTA 120
DB 61 GTTTTAGCTTTAATTCTAAGAGAGATGACTAATGAGCAGAAATGCAAAAACTA 120
QY 121 AAAAAATTTGGCAAGATGACTGAGCAAAAATCTAATATAGAACAGTACAGAAA 180
DB 121 AAAAAATTTGGCAAGATGACTGAGCAAAAATCTAATATAGAACAGTACAGAAA 180
QY 181 CTTAAGGTTTTTGGAGAGTGGAAAGAGATGATACAAAATGCAAGAACTGAAAGCCAT 240
DB 181 CTTAAGGTTTTTGGAGAGTGGAAAGAGATGATACAAAATGCAAGAACTGAAAGCCAT 240
QY 241 ATTGAGAAAAATGTACTACATCAAGAGAAAACTTAAAGACAAATTTAAAAAAATTT 300
DB 241 ATTGAGAAAAATGTACTACATCAAGAGAAAACTTAAAGACAAATTTAAAAAAATTT 300
QY 301 CAGATTATACGAGATAGGATTTGCAAGAGATGAAACAATGCTATTTTGGAGGGA 360
DB 301 CAGATTATACGAGATAGGATTTGCAAGAGATGAAACAATGCTATTTTGGAGGGA 360
QY 361 GATGTTCAAAAAGAACTTAAAGATGATGCAATGCTTTGAGAAATAGGCTATCAAAAG 420
DB 361 GATGTTCAAAAAGAACTTAAAGATGATGCAATGCTTTGAGAAATAGGCTATCAAAAG 420
QY 421 AAACGTGATAAAGTTGCGAGAGAGATTTCTTTTAAAGACATTCGTAGCGATCTTAATGA 480
DB 421 AAACGTGATAAAGTTGCGAGAGAGATTTCTTTTAAAGACATTCGTAGCGATCTTAATGA 480
QY 481 TCAGTCATATGTGAAAAAACTTAAAGAGATTTGCCCTGTATGCGGAGGAAAGTAT 540
DB 481 TCAGTCATATGTGAAAAAACTTAAAGAGATTTGCCCTGTATGCGGAGGAAAGTAT 540
QY 541 GAGTTAACAAACTGTGTCTGAACAGAAAGAGACATGATGAAATATTTTAAATGAAAA 600
DB 541 GAGTTAACAAACTGTGTGTCTGAACAGAAAGAGACATGATGAAATATTTTAAATGAAAA 600
QY 601 GATTAAGAGTGGGATCTCTTAAAAAGATGTTTCAAGCAGCCTAGGAAGTTTAAAAAA 660
DB 601 GATTAAGAGTGGGATCTCTTAAAAAGATGTTTCAAGCAGCCTAGGAAGTTTAAAAAA 660
QY 661 GAAACATGCTTTGAATTAATCTGAAACAATGCTATTTTCAATGGAAATTTGCGAGACAG 720
DB 661 GAAACATGCTTTGAATTAATCTGAAACAATGCTATTTTCAATGGAAATTTGCGAGACAG 720
QY 721 GATTAATTAATTAATGATTAATGAGAGAGAAATGCAAGAACAAATATGCTTAATATG 780
DB 721 GATTAATTAATTAATGATTAATGAGAGAGAAATGCAAGAACAAATATGCTTAATATG 780
QY 781 CCACCAAGGACCCGATTTTGAATCAATAGGCCAGAGGCTACAAATAGCAGAGATATAGGG 840
DB 781 CCACCAAGGACCCGATTTTGAATCAATAGGCCAGAGGCTACAAATAGCAGAGATATAGGG 840
QY 841 CTGGAAGAGTTTTTATTAAGAGATGAGAGAGATGAGATTTTATTTGGAAGAAATCATCTA 900
DB 841 CTGGAAGAGTTTTTATTAAGAGATGAGAGAGATGAGATTTTATTTGGAAGAAATCATCTA 900
QY 901 AGAGATGCGACAGCTTTGTTGGCATTTGTTGATCCAAAGATTTCAATCTTAAAAAAAGAGC 960
DB 901 AGAGATGCGACAGCTTTGTTGGCATTTGTTGATCCAAAGATTTCAATCTTAAAAAAAGAGC 960
QY 961 GACAAAGAGAAATGCGAAGAGAGCCCTTCAAAAAAGCTGCAAAATATCTCATATAGATAG 1020
DB 961 GACAAAGAGAAATGCGAAGAGAGCCCTTCAAAAAAGCTGCAAAATATCTCATATAGATAG 1020
QY 1021 GCTTTAGAAAATTTATGTAAGAAAAATGCTTAAATGATGAGAACGAAAAATGTGAA 1080

DB 1021 GCTTTAGAAAATTTATGTAAGAAAAATGCTTAAATGATGATGAGAACGAAAAATGTGAA 1080
QY 1081 GAATTGCAAAATGATATTTAACAAAACCTTGCAAAATTTTCACTTCAAAAAGTCACTAATAT 1140
DB 1081 GAATTGCAAAATGATATTTAACAAAACCTTGCAAAATTTTCACTTCAAAAAGTCACTAATAT 1140
QY 1141 GGTCTTTTGTATCCAAACAAAAGAAATATGAAATGTTGGATGCGAAGGTTGCCAACA 1200
DB 1141 GGTCTTTTGTATCCAAACAAAAGAAATATGAAATGTTGGATGCGAAGGTTGCCAACA 1200
QY 1201 TTTCTTGAACAGAAAGATTTGGCGAAATTTGAGAGCTATTTGTTCTATTTTGAAAAAAA 1260
DB 1201 TTTCTTGAACAGAAAGATTTGGCGAAATTTGAGAGCTATTTGTTCTATTTTGAAAAAAA 1260
QY 1261 TGTCCAGATGAGAAAAATGCAATGTAATAATGTAAGAGCAATGTTCAAAAAGAGACTT 1320
DB 1261 TGTCCAGATGAGAAAAATGCAATGTAATAATGTAAGAGCAATGTTCAAAAAGAGACTT 1320
QY 1321 GATCAGCGGCAAAATTAAGTCTGCAAGAAAAATGCGAGAGATGTTACATGTTCAAAAC 1380
DB 1321 GATCAGCGGCAAAATTAAGTCTGCAAGAAAAATGCGAGAGATGTTACATGTTCAAAAC 1380
QY 1381 AAAAGCTGGCTTGAAGATTTTCAACAAAGATTAATGTAAGATGAGAACTGAAAAAA 1440
DB 1381 AAAAGCTGGCTTGAAGATTTTCAACAAAGATTAATGTAAGATGAGAACTGAAAAAA 1440
QY 1441 GAAAAACAAGAGAAATTTCTCAACAGATGATATTTATTTCTGTGTATGACCCAGCAAAA 1500
DB 1441 GAAAAACAAGAGAAATTTCTCAACAGATGATATTTATTTATTTCTGTGTATGACCCAGCAAAA 1500
QY 1501 GCAGCCGGTGTCTTACACATGATCTTGAAATGAAACATATCTTTTAAACAAACTG 1560
DB 1501 GCAGCCGGTGTCTTACACATGATCTTGAAATGAAACATATCTTTTAAACAAACTG 1560
QY 1561 GATCAAAAAGCAGATTTTCCGACAGATTAATAATGCAAGAAATTTGGGAGAAAGTCCAA 1620
DB 1561 GATCAAAAAGCAGATTTTCCGACAGATTAATAATGCAAGAAATTTGGGAGAAAGTCCAA 1620
QY 1621 GATTTAGAGAGATTTCAAAAAGATTAACATGAGCATCTGATACCTGAGACAGCAATGC 1680
DB 1621 GATTTAGAGAGATTTCAAAAAGATTAACATGAGCATCTGATACCTGAGACAGCAATGC 1680
QY 1681 AATGCTTTGGGACTACAGAAATTTTAAAGCAGGTTTATGATGATGAAACACAAAGATATCT 1740
DB 1681 AATGCTTTGGGACTACAGAAATTTTAAAGCAGGTTTATGATGATGATGAAACACAAAGATATCT 1740
QY 1741 TTGAAGAACAAGAAATGTTAATAATTAATCTTAAAGAAAGTGAATTAATGCTATGAG 1800
DB 1741 TTGAAGAACAAGAAATGTTAATAATTAATCTTAAAGAAAGTGAATTAATGCTATGAG 1800
QY 1801 AGAGAGATGACGTTTCTCTTTGTATGTCTTCCAAAACGCTACGTTGAGCTGATG 1860
DB 1801 AGAGAGATGACGTTTCTCTTTGTATGTCTTCCAAAACGCTACGTTGAGCTGATG 1860
QY 1861 GTAAAGAAGCTGAAGAAGAGGATGAGATTTCAAAAAAAATATATTAAGCTTCAATATAT 1920
DB 1861 GTAAAGAAGCTGAAGAAGAGGATGAGATTTCAAAAAAAATATATTAAGCTTCAATATAT 1920
QY 1921 ATTGAATTTCTTGAATAATTAATCAATAATTAACAACCTGGAAGAAATTTGCCCTCT 1980
DB 1921 ATTGAATTTCTTGAATAATTAATCAATAATTAACAACCTGGAAGAAATTTGCCCTCT 1980
QY 1981 TGGCATACGATTTCAATAGATTTTCACTAATTTGTCAGAGGCTTTTCAAGAAAGAGATAGT 2040
DB 1981 TGGCATACGATTTCAATAGATTTTCACTAATTTGTCAGAGGCTTTTCAAGAAAGAGATAGT 2040
QY 2041 TGTACAAAATATCAAGAAAGATCTGAGGCGTTCTATTAAGAAAGAGGCTTGTGAAGATCT 2100
DB 2101 TGTACAAAATATCAAGAAAGATCTGAGGCGTTCTATTAAGAAAGAGGCTTGTGAAGATCT 2100

Qy	721	GATGTAATTAATGATATTGGAATTTGGAGGAAATGCCAGAAACAAATATTCGTATATG	780
Db	718	GATGTAATTAATGATATTGGAATTTGGAGGAAATGCCAGAAACAAATATTCGTATATTA	777
Qy	781	CCACACGAGACCCGATTTTGATCCAACTAGAGCCAGAGGCTCAATATGACAGAGATATAGG	840
Db	778	CCACACGAGACCCGATTTTGATCCAACTAGAGCCAGAGGCTCACTAGACAGAGACATAGC	837
Qy	841	CTGGAGAGCTTTATTAAGAAAGTAGAGAGATGAGGTTTATTTAGGAAAGATCATCTA	900
Db	838	CTGGATGAGCTTTATTAAGAAAGTAGAGATGAGGTTTATTTAGGAAACATCATTTA	897
Qy	901	AGAGTGGGACAGCTTTGTTGGCATTTGTTG---ATCCAGATCTCTAGTCTTAAAAAAA	957
Db	898	AGAGTGGGACAGCTTTATTTGACCTTTGTTGTTAGAGAGATGATACAGAAAAATATAT	957
Qy	958	GACGCAAAAGAAATCGAAGAGCCCTTCAAAAAAGCTCAAAAAATCCTCATGACAT	1017
Db	958	AATATCGAGAAAAATCGAATATGATTCCTGAGATATTAATGCAAAAACCTCAGACGAT	1017
Qy	1018	GAGGCTTTAGAAAGTTATGTAAAGAAAAATGTTTAAAGTATGATGAGAAAGAAAAATGT	1077
Db	1018	GAACTCTAAAAAATTTATGTATATTAATATATGCTTAATGCAATGAAAAAGAAAAATGC	1077
Qy	1078	GAGAAATTCGAAAAATGATTTAAACAAACTGCAAAATTTTCACTTCAAAAGCTCATAT	1137
Db	1078	AAAGATTTAGAGAAATATTTAAAAAACATGCAAAACCTCAACCAACGATTTCTTAA	1137
Qy	1138	AATGCTCTTTTGTATCCAAACAAAAGAAATATGAAATGTTGGATGGAGGGGTGCCA	1197
Db	1138	AAACATCTTTTGTATCC-----AAATGATTAATTTGTTAGTGAAGAAAACTGCCA	1188
Qy	1198	ACATTTCTTAGACAAGAGATTGTGCGAATTTGAGTCCATTTGTTCTATTTTGA AAA	1257
Db	1189	ACATTTCTTACTAATGAAAGCTGTGCAAGATTGGAACTTATTTGTTTATCTACAAAAA	1248
Qy	1258	AAATGTCAGATGAGAAAAATGCAATGTAATAATATAAGACCAATGTTACAAAAAGAGA	1317
Db	1249	GCTGTCCAAATGCCAAAGAAAGATGTTGAAATTTGAGGGCAGCGTCTTATAGAGGG	1308
Qy	1318	CTTGAATGACGGGCAAAATAAATGCTGCAAGAAAAATATGCGAAGAAATGTTACATGGTCA	1377
Db	1309	CTTGAATGACGGGCAAAATAAATGCTGCAAGAAAAATATGCGTGGTTATTAAGTGGTCA	1368
Qy	1378	AACAAAGCTGGCTTGA AAAAGTTTCAACAGAAATTTAGTAAAGTATGTGAGAAA	1437
Db	1369	AATCAAGATGGCTTAAAGAGATTTCAACAAAGATTTGTAATAATATGTAAGAGAGCT--A	1425
Qy	1438	AAAGAAAAACAAGAGATTTCTCAAAAGATGAAATTTATTTGTGTGTACAGCCAGCA	1497
Db	1426	AAAGAAATTAAGAGAGTTTCCAAACGATGAAATTTTCTGTGTGTACAGCCAGCA	1485
Qy	1498	AAAGACGCGGTTGCTTACACATGATCTTTCGAATGAAAACTATCTTTTTCAGACA CA	1557
Db	1486	AAAGGTGACATTTATCTTACACAGATCATCAATGAGGGTTATCTTTTTCAGACA CA	1545
Qy	1558	CTGATGCAAAAGGAGATTTCCGCAAGATTA AAAATTTGCAAGGAATTTGGGGAGAAATGC	1617
Db	1546	TTGGATCAAAAAGAGAGATTTTCGACAGATTAAGATGCAAGGAATTTAGGAAAAATGC	1605
Qy	1618	CAAGATTTAGAGAGATTTCAAAAAGAAATTCATGCGCATGTATCTACTGAGAGCA	1677
Db	1606	CAAGATTTAGAGAAAGATTTCAAAAAGAAATTTKATGCGCATGTATACGCTGAGAGCA	1665
Qy	1678	TGCAATGCTTTGGGACTACAGAAATTTTAAAGCAGGTTTATTTGATGAACA CAAGAT	1737
Db	1666	TGCAATGCTTTGGGAGACTACAGAAATTTTAAAGCAGGTTTATTTGATGAACA CAAGAT	1725
Qy	1738	ACTTTGAAGAACAAGAAAGTTGTGTAAATTAACCTA AAAAGAGTATATAAGTCT	1797
Db	1726	ACTTTGAAGAACAAGAAAGTTGTGTAAATTAACCTA AAAAGAGTATATAAGTCT	1785


```
QY 61 GTTTAGCTTTAATTCTPAAGAGATGSACTAAGTGAGCAGAGTGCAGAAAAAACTA 120
D 61 GTTTAGCTTTGATTTTAAAAAATGATTTAGAGATACAAAAATGCAGAACTAAGTG 120
QY 121 AAAAAATATGCGAAGATTGACTGAGCAAACTAAATATATGAAACAAGTACAGAAAA 180
D 121 GAAAGATATTTGCAAAACATTAACAATGCAAGATTTAAATCCAGAAAAAGTTCAAGAAAA 180
QY 181 CTTAAAGCTTTTCCGAGATGAGAAAAAGATACAAAAATGCAAAAGAACTGAAAGCCAA 240
D 181 TTTAAAGATTTCTGTGATACGGGAAAGAAATGAAAAATGTCAAGATCTTAAAAACAA 240
QY 241 ATTGAGAAAAATGTACTACATCAACAGAAAACTTAAAGAAAGAACTTAAAAAAATTT 300
D 241 GTCAATCAAAAATGATTTAAATTTCCAGGAAAACTTCAAAAGCTGCTGAAAAAAATTT 300
QY 301 CAGATTAACGAGTAAGAGATTGCAAGAGAAATGACAAATGCTTATTTTGGAGGGA 360
D 301 TCAGAAATTAACAGATGAGAGATTGCAAAAAAGAAATGACAAATGCTTATTTTGGAGGGA 360
QY 361 GTATGTTCAAAAGAACTTAAAGATGATGCAATCTTTGAGAAATTAAGTGTATCAAAAG 420
D 361 GCATGTCCAAAGAACTTAAAGATGATGCAATTAATTAAGAAATTAATCTGTTATCAAAA 420
QY 421 AAACGTGATTAAGTTGCGAGAAAGTTCTTTAAGAGCACTTCGAGAGATCTTAAATGGA 480
D 421 GAACGGAACATGTGGCAGAGAAAGTTCTTTGAGGCGCTTCGTGATCTCAATGAA 480
QY 481 TCAGTCAATGTGAAAAAACTTAAAGAGATTGCGCTGTCAATGAGGAGGAAAGTAT 540
D 481 ACAAAAGCATGTGAAAAAAAGCTGAAGAGATTGGCCGAAATTAAGAAAGAAAGCGAT 540
QY 541 GAGTTAACAACTTGTGTCTGACACAGAAAGACATGTAAGATATTTTAATGAAAA 600
D 541 GAAATTAACGAGCTTGTCTTATCAAAAAACAATGCGTAAGTCTTGTAAACAAAAAGA 600
QY 601 GATAAGAGTGCAGACTCTTAAACAGATGTTTACAGAGCACTAGGAAGTTTAAA--- 657
D 601 AAAAGTAATGTGATACCTTTGAAAAAGAGATTGAAGAGCACTTAAAGAGATGAATTTG 660
QY 658 AAAGAAACATGTCTTGAATTAATCGAACAAATGCTAATTTTACATTTGAAATTTGCGAG-- 715
D 661 CGAGAAAAATGTCTAATTAATTAATCTTACAGAGATGTTTCAAGAGGGAATGTAAGGA 720
QY 716 -----ACGACATATTAATTAATGT 735
D 721 GACAAATCAAGTCAATTAATTAATTAAGATGCAAGAAATATGTACAGAGTGT 780
QY 736 ATTGAATTTGGAGGAAATGCGCAAGAACAAATATTTGTTATATGCGACCGAGCCGAT 795
D 781 GATGAATTTAGCAAGAAAGTGTGAAAGAAATATGTTATATATGATCCAGAGTCCGAT 840
QY 796 TTTGATCCAACTAGAGCCAGAGCTACATATAGCAGAGATATAGGGCTGGAAGATTTTAT 855
D 841 TTGATCCAACTAGAGCCAGAGCTACATAGCAGAGCATATAGGGCTGGAAGATTTTAT 900
QY 856 AAGAAGGTAGAGAGATGAGTTTATTTGAAAAGAAATCATCTTAAGAGATGCGACGCT 915
D 901 AAGAGGCGAGAGAGATGAGTTTGTGTGAAAGACAAATGTATGAGATGCAACGCT 960
QY 916 TTGTTGGATTTGATGCAAGATTTCTAGCTTAAAAAAAAGCAAAAGAGAAATGC 975
D 961 TTGTTGGACTACT-----TCTTAAGAAACCTTTAAAAAAGAAAGATGT 1005
QY 976 GAAGAAGCCCTTCAAAAAAGCTGCAAAATCTCATGAAACATGAGGCTTTTGAAGTTTA 1035
D 1006 ATAAAGCCCTTAAAAAAGCTGCGAAAACTCATGAAACATGAGGCTTTTGAAGATCTA 1065
QY 1036 TGTAAAGAAAAATGTTTAAGTAAATGATGAAAGAAAAATGTAAGAAATGCAAAATGAT 1095
D 1066 TGTAAAGAAAAATTAACCAATGATGATGAAAGAAAAATGTGATGAATGAAGAAAAAGAT 1125
QY 1096 ATTAACAAAATCTGCAGAAATTTTCACTTCAAAAGTCACTAATAATGCTTTTGTATCCA 1155
D 1126 GTTAACAAAATCTGTAACAGCTTACATCAACATTTCTTAAAAACCGTCTTACATTTCA 1185
QY 1156 ACAAAAGAAATATGAAATTTGATGATGGAAGGGTTGCCACATTTCTTGAACAAGAA 1215
D 1186 CC-----TGTAGAAATTTGGGAAATGGGAAAAATTAACCGACATTTCTTATGATGAA 1236
QY 1216 GATTGCGAAATTTGAGTCTTATTTGTTCTAATTTTGAIAAAAAATGTCAGATGAGAA 1275
D 1237 GATTGCGAAATCTAGAAATCTTATTTGTTTATTAAGAAACTTGTCCAGATGTCAAA 1296
QY 1276 AATGATGTAAAAATATATAGAGCAACATGTTACAAAAGAGACTGTATGACAGGGCAAT 1335
D 1297 GAAGCTGTATGATGATGAGAGGCGGTGTTATTAAGAGAGGGCTGTATGACAGGGCAAC 1356
QY 1336 AAGTGTGCAAGAAATATGCGAGAAATGTATACATGTTCAACAAAGCTGCTTGA 1395
D 1357 AGTGTGTTGCAAAAAATATAGCGAGGTTATTTGATGCTCAATTAAGATTTGCTTAAG 1416
QY 1396 AAGTTCAACAGAAATTAATGTAAGATATGTAGAGAACTGAAAAAGAAACAAAGAGT 1455
D 1417 AAATTTCAACAGAAATTAAGCAAAAGTATGTAGAACTG---AAGGAAATTAAGAGAGT 1473
QY 1456 TTCTCAACAGATGATTAATTAATCTGTGTGTACAGCCAGCAAAAGAGCCGGTGTCTT 1515
D 1474 TTCTCGAACAGATGATTTGTTGTTCTGTGTATACACAGAAAGGACGACGATTACTT 1533
QY 1516 ACACATGATCTTGAATGAAAACTATCTTTTACGACACACATGATCAAAAGCGAGAT 1575
D 1534 ACACATGATCATCAATGAAGAGTATCTTTTACGACACACATGATCAAAAGAGAGAT 1593
QY 1576 TTCCCGACAGATTAATAATGCAAGGAATTTGGGGAGAAAGTCCAAAGTTTAAGAGAGAT 1635
D 1594 TTTCGACAGATTAAGCTGCAAGAAATTAAGGAGAAATGCAAGATTTTAAGAGAGAT 1653
QY 1636 TCAAAAGAAATTAATCATGCGCATGTATACACTGAGAGCAATGCAATTCGTTGGGACT 1695
D 1654 TCAAAAGAAATTAATCATGCGCATGTATACACTGAGAGCAATGCAATTCGTTGGGACT 1713
QY 1696 ACAGAAATTTTAAAGCAGGTTTATTTGATGAAACAAAGATCTTTGAAAGACCAAGAA 1755
D 1714 ACAGAAATTTTAAAGCAGATTTTATTTGATGAAACAAAGATCTTTGAAAGACCAAGAA 1773
QY 1756 AGTTGTGTAATAATCCTAAGAAAGAAAGTAAATTAAGTCTGAAGAGAGATGACGT 1815
D 1774 AACTGTCAAAATTAATTTAAAGAAAGAAATGCAATTAATGCTGAGAGGGTGAATGCT 1833
QY 1816 TTCTCTTTGTATGTCTTCAAAAGCCTACGTGAGCTGATGTAAGAAAGAGCTGAA 1875
D 1834 TTTTCTTTGTATGTCTTCAAAAGCCTACGTGAGCTGATGTAAGAAAGAGCTGAA 1893
QY 1876 GACAGGTGTGAAGTATTAAGAAAGAAATATTAAGCTTCAATATTAATTAATTTCTTGA 1935
D 1894 GATAGGTGCAAAATATTCGAAGAAATATGCAAGCATCAATATTAATTAATTTCTTAAA 1953
QY 1936 AATTAATCAATTAATAAATTAACACATGGAAGAAATTTGCTCCCTTGGCAATACCTATTC 1995
D 1954 AAAAAATCAATTAATAAAGCAATGACAGCAAAATTTGTCCCTCATGACATCCATTCG 2013
QY 1996 AATGATTTTCACTAATTTGCGAGGTCTTAAGAAAGAGAAATAGT---TGTACAAAATC 2052
D 2014 GATAGATTTTATCCCATTTGTCTGATCTTAAGAAAGAAACTTCTGTCAAAATCTT 2073
QY 2053 AAGAGCATGATGAGCGGTTCTATTAAGAAAGAGCTTGAAGATGCTTCAAAAGTGA 2112
D 2074 AAAAAATATGCGAACATTTCTACAAAAGAAAGTTTAAAGATGCTTAAAGTGA 2133
QY 2113 CTTCAGAGAAATGATGATGAAATCTTAATTAATGAAAGCTGCAATTTGAAGATTTGA 2172
D 2134 CTTCAGAGAAATTTAATTAATTAATTAATTAATGAAAGCTGCAATTTGAAGATTTGA 2193
QY 2173 GTAGCGGAAACGTAAATTAATGCGTCAATGATGCTTAATGCAAGCTAACACCAAGAT 2232
```

Db 2194 GTATTGAAAGAGCTAAATATATCGTCATACAGCAGTTTATGTAAAGATTAATCCAAAAGT 2253
Oy 2233 AACTCTGGAAAG---AGTGATGAGATGCTAGAAAAGAACTGTGAGAAATTAAGTGA 2289
Db 2254 AAAAATAAAAAGCCGATAATAAAATGTTAAGAAAAGAACTGTCTAATAATTAAGTGA 2313
Oy 2290 GAAATGGAAGAACAGTGCAGAACATTTACCAAGAAATTAAGCAACCCGACAGCTGATCTA 2349
Db 2314 GAGGTGGAACACAGCAATGCAAAATATTACCAAGAAATTAAGCAAGCTGGAATAAGTCTA 2373
Oy 2350 AAAAAAGATTATTAAGCATATATAGAACTTAAGAAACGTGCAGAGAAAGCAATGAACAAG 2409
Db 2374 AAAAAAGATTATTAAGCATATATAGAACTTAAGAAAGGCAAAAAGCAATGAACAAG 2433
Oy 2410 TCCAGCTCTTTTGTGTCACTCATTAAGAAAAGCAAAAGTAATGTATCAAAAAGTAATGAC 2469
Db 2434 TCCAGCTCTTTTGTGTCACTCATTAAGAAAAGCAAAAGTAATGTATCAAAAAGTAATGAC 2493
Oy 2470 AAAAAAGATTATTAAGCATATATAGAACTTAAGAAACGTGCAGAGAAAGCAATGAACAAG 2529
Db 2494 AAAAAAGATTATTAAGCATATATAGAACTTAAGAAAGGCAAAAAGCAATGAACAAG 2553
Oy 2530 AATCTACGGAAGGAGATTAAAGATGTATCCGTAAAGCAAAATTAAGCAATTAAGCAATTTGAT 2589
Db 2554 AATCTACGGAAGGAGATTAAAGATGTATCCGTAAAGCAAAATTAAGCAATTTGAT 2613
Oy 2590 TTGGCAGCAGAAAGTATTTGGAAGATATGAGATTTGAAGAAAGATGTATTAATTTGAA 2649
Db 2614 TTGGCAGCAGAAAGTATTTGGAAGATATGAGATTTGAAGAAAGATGTATTAATTTGAA 2673
Oy 2650 TCAGATTGCAAAATTAAGAGAGATTGCAAAAGCTTAAGAAAGATATGCAAAAAGTAAT 2709
Db 2674 TCAGATTGCGGAGATTAAAGACGATTCGATGCTTTAAAGAAAGTGTGCAAAAGATTGAG 2733
Oy 2710 AAGGCTTGCGGAATCTGAAGCTCTGAGAGTGAAGCCGCAAGAAAGCTGACAGAAAGT 2769
Db 2734 AAGACATGTCAGATCTGAAGCCTCTGAGAGTGAAGCTGCAATGAATATGTCACAGAAAGC 2793
Oy 2770 ACAACGACAACTTAACAACAACAACAACCGTTGCGATCCGAAAGCAACGAAATGCAAA 2829
Db 2794 ACAACGACAACTTAACAACAACAACAACCGTTGCGATCCGAAAGCAACGAAATGCAAA 2853
Oy 2830 TCCTTAACAGCAACAGACATAGGTTTACAGACATGACACACACACACAGCTCTACT 2889
Db 2854 TCCTTAACAGCAACAGATCATAGGTTTACAGACATGACACACACACAGCTCTACT 2913
Oy 2890 ATCACAATCTACATCATCAATCAAAAATTAATGACATGACATGACAGCGATGCAACCAACC 2949
Db 2914 ATCACAATCTACATCATCAATCAAAAATTAATGACATGACATGACAGCGATGCAACCAACC 2973
Oy 2950 AAGTGTACGACAGAGAGAAAGATGATGACAGAGAGCTGAAACCGAGTGAAGGGCTGAGG 3009
Db 2974 AAGTGTACGACAGG-----GGATGAAGAGAGAGAGCTGAAACCGAGTGAAGGGATTTGAAG 3027
Oy 3010 ATGAGTGGGTGGAATGTGTGAGGGGGGTGATGATGACAAATGTTTCTTCATGATTT 3069
Db 3028 ATGAGTGGGTGGAAGGTGATGAGGGGGGTGATGATGACAAATGTTTCTTCATGATTT 3087
Oy 3070 TAG 3072
Db 3088 TAG 3090

; CURRENT APPLICATION NUMBER: US/10/654,416
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/09/762,724
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3081
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3030)
US-10-654-416-9
Query Match 52.9%; Score 1624.8; DB 19; Length 3081;
Best Local Similarity 72.8%; Pred. No. 2.9e-311;
Matches 2275; Conservative 0; Mismatches 752; Indels 99; Gaps 10;
Oy 1 ATGGCGCGGGGCTCAAGCCGACG-----GCAGAGGGAGACAGAAATAGCATTTGAT 51
Db 1 ATGGCGCGGGGCTCAAGCCGACGCTGCAGAAACGATCAGGGGCTAGTATATGATGATGT 60
Oy 52 GAGGAACATGTTTAAAGCTTTAATCTAAAGAGATGAGCTAAGTGACAGAAATGCAAA 111
Db 61 GAAAGAAATCTTTTGGCTTAATTTGCA---CGAAGAAATTAATGATTAATGAAATGCAAA 117
Oy 112 AAAAAAATTAATAATATTTGCAAGAAATTTGACTGAGCAAAATCTAATATATGAAACAAGTA 171
Db 118 AAGAGATTAGAAAAATATTTGTAAGACATTTAAGATGCAAGAAATTTAAACCAAAAAAGTT 177
Oy 172 CACAGAAATCTTAAGGTTTTTGGAGAGATGGAAGAGATGATACAAAAATGAAAGAACTG 231
Db 178 CACAAAAATCTTAAGAGTTTGTGAATAATTAAGAAAGCAAGTTCAAAATGCAAAAGAACTG 237
Oy 232 AAACCATATTTAGAAAAAATATGTAATCAATCAAAAGAAACCTTAAGAAAGCAATTTAA 291
Db 238 AAGAAATCACTCACTCAAAATATGTAAGCAATCAAAAGAAACCTTAAGAAAGCAATTTAA 297
Oy 292 AAAAAAATTCAGATTATTAAGGATTAAGGATTTGCAAGAAATGAAACAACAATGCTTATTT 351
Db 298 AAAAAAATTCAGATTATTAAGGATTAAGGATTTGCAAGAAATGAAACAACAATGCTTATTT 357
Oy 352 TTGAGAGAGATATTTTCAAAAAGCTTAAGATTAAGTATGCAATCTTTGAGAAATTAAGTGC 411
Db 358 TTGAGAGAGATATTTTCAAAAAGCTTAAGATTAAGTATGCAATCTTTGAGAAATTAAGTGC 417
Oy 412 TATCAAAAAGAAATGATTAAGATTTGCGGAAGAAATTTCTTTTAAGAGCACTTGTAGCAT 471
Db 418 TATCAAAAAGAAATGATTAAGATTTGCGGAAGAAATTTCTTTTAAGAGCACTTGTAGCAT 477
Oy 472 CTTAATGATCAGTCATATGTGAAAAAATTAAGAGATTTGCTGTATGAGGGAGG 531
Db 478 CTTAATGATCAGTCATATGTGAAAAAATTAAGAGATTTGCTGTATGAGGGAGG 537
Oy 532 GAAAGTATGAGTTTAAACAATTTGTGTCTGAACCAAGAAAGAGATGTAAGAAATTTTAA 591
Db 538 GAAAGTATGAGTTTAAACAATTTGTGTCTTTTATCAAAAAATGATGTAAGAAATTTTAA 597
Oy 592 ATTGAAAAAGATTAAGAGTGGGATCTTTTAAACAGATGTTTGAAGCAGCACTAG----- 646
Db 598 TTGAAAAAACAATAATATGATGCTTTTAAACAGATGTTTAAACGACGACTTGAAGAG 657
Oy 647 -GAAGTTTAAAAAAGAAACATGCTTGAATTAATCTGAAACATGCTATTTTAACTTGA 705
Db 658 AAGATGAGTTACAGAGAAATGTTTACCACTGCTGAACATGCTATTTTATGAGGG 717
Oy 706 AATTCGAGG-----ACGAGATTAATTT 729

RESULT 5
US-10-654-416-9
; Sequence 9, Application US/10654416
; Publication No. US20040043409A1
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; TITLE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii
; FILE REFERENCE: 4239-58054

D 718 AATTGAGATATATCATAAATGTATTAATCATCCGAGACTGTATGATATTTGCCA 777
Q 730 AATGTATGTAATGGAGAGAAATGCCAGACAAATAATGCTTATATGCCACGAGA 789
D 778 GTGTGATATCATTTGGCAGTGAATAATGTAGAAAATTAATTTATATACATCCGGGA 837
Q 790 CCCGATTTTTCATCACTAGCCAGAGGCTCAATAGCAGAGATATAGGCTGGAAG 849
D 838 TCCGATTTTCAATCAATCAATAGCTCAAGCTTCTAGCAGAGACATAGCAGTGGAAAG 897
Q 850 TTTTATTAAGAGTAGAGAGAGATGAGTTTATTTGAAAAGATCATTTAGAGATCG 909
D 898 CTTTATTAATAAGAGAGAGAGAGATGAGTTTATTTGAAAAGATCATTTAGAGATCG 957
Q 910 AAGCTTTTGTGGATGTTGATCCAGATTTCTAGTCTTAAATAAAGAGCAGAAAGAG 969
D 958 ACTGCTTACTGGGCTTTGATTTCAAAATCTAGATCTTAAGATCAAGTGGTAAAG- 1015
Q 970 AATGCGAGAGAGCTTTCAAAAAGCTGCAAAAATCCTCATGAACTAGAGCTTGAAG 1029
D 1016 -AATGCGAGAGAGCTTTCAAAAAGCTGCAAAAATCCTCATGAACTAGAGCTTGAAG 1074
Q 1030 AGTTATGTAAAGAAAATGTTTAAATGATGAGAAAGAAAATGTAAGAAATTCGA 1089
D 1075 GATTTTGTATCAAAAATGATGAGTGTCAAAAATGAAATGAAAGTGAAGATAGAG 1134
Q 1090 AATGATATTAACAAACTTGCAGAAATTTTCTCAAAAGCTCAATTAATCGCTTTT 1149
D 1135 AAGAGATTAGCAACAGTATCTAAATATCTTTTAAATAAAGAAATTAACACCTCT 1194
Q 1150 GATCAACAAAAGAGAAATGAATGATGAGTGGAGAGGTTGCCAACTTTCTTAGC 1209
D 1195 -----GATCCGAGAGAGCTTTCAAAAGCTCAATTAATCGCTTTT 1242
Q 1210 AAGCAGATGTGTGCAAAATGAGATGCTTATTTCTATTTTGAATAAATAATGCTCAGAT 1269
D 1243 GACATATCTGCAACAGTATGAGTCAAGCTGTTTATTTT-----AAAAGCTCAAGCAGCT 1299
Q 1270 GGAGAAAATGATGTAATAAATATAGAGCAATGTTTACAAAGAGAGCTTGAATGACCG 1329
D 1300 CTGACAAAGAGATGTATTAATCTAGAGCAGATGTTTAAAGAGAGCTTGAAGCAAA 1359
Q 1330 GCAATATAAGTCTGCAAGAAAATATGCGAGAGATGTTACATGGTTCAACAAAGCTG 1389
D 1360 GCTAATAGAGATGCGAGAAAATATGAGAGCTGTTCTATGTTTCAAGGAAAGATG 1419
Q 1390 CTGAAAAGTTCACACAGAAATAGTAAAGATGAGAACTGAAAAGAAAAGAAACAA 1449
D 1420 TTTTAAAGAACTATAGAAAATAATGAGAGATGTTCCGAACTTTAAACA----- 1470
Q 1450 GGAAGTTCTCAACGATGATTTATTTATTTGTGTGTACAGCCAGCAAAAAGCCCG 1509
D 1471 -----ACAGCAGATGAGTTGTTTGTCTATGATGATCCACTTAAAGCAGTCA 1521
Q 1510 TTGCTTACATGATCTTCGAATGAAAATCTATCTTTTACGACAACTGATCAAAAG 1569
D 1522 ATACTTTCAGCTGATATTCAGAGCAAGCAGCTTTTTCGAGAAAATTTGATCAAAAG 1581
Q 1570 CGAGATTTCCGACAGATTAATAATGCAAGAAATGGGAGAGAAAGTGCAGATTTTGA 1629
D 1582 CGAGACTTTCCAGACAGAAAGATTTGCAAGAAATTTAGAGAAAGATGTAAGCTTTAGGG 1641
Q 1630 GAGATTTCAAAAGAAATTAATCATGCGCATGTATACATGAGAGCAATGCAATCGCTT 1689
D 1642 AAGATTTCAATCAGATTAATGATGCGCATGTATACGCTTAAACAAAGTGTGATCGCTT 1701
Q 1690 GGAATCTACAGAAATTTTAAAGCAGATTTTATTTGATGAAACAAAGATTAATGAAAGAC 1749
D 1702 GGGATCTACAGAAATTTTAAAGCAGATTTTATTTGATGAAACAAAGATTAATGAAAGAC 1761
Q 1750 CAAAGAAATGTTGTAATAATTAATCAATTAAGAAAGATGTAATAATGTTGTAAGAGAGAT 1809
D 1762 CATGAAAATCTGTAGAAATTAATTAAGAAAGAAATGTATTAATGTTGTAAGAGAGAT 1821

Q 1810 GACGTTTCTTTTGTATGTCTTTCAAAAGCTACGTGTAGCTGATGTAAGAAAGAC 1869
D 1822 GATCGTTTCTTTTGTATGTCTTTCAAAAGCTACGTGTAGCTGATGTAAGAAAGAC 1881
Q 1870 GTGAAAGACAGGTGTGAAGTATTCAAAATAATTAATAAGCTTCAATATTAATTTGAATTT 1929
D 1882 GTGAAAGACAGGTGTGAAGTATTCAAAATAATTAATAAGCTTCAATATTAATTTGAATTT 1941
Q 1930 CTTGAAAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTA 1989
D 1942 CTTGAAAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTA 2001
Q 1990 TATTCATATGATTTTCACTTAATGTCAGCTTTCAGAAAGAGATAGT---TGTACA 2046
D 2002 TACTGATATGATTTTATCCCAATGTCATGATCTTAAAGAAAGAAACCTTCTGTCA 2061
Q 2047 AATCTTAATAATTAATGCGAATCTCAAAAAGAGGTTTAAAGATGCTTTAA 2106
D 2062 GTAGAGCTTCAAGGAAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2121
Q 2107 GTAGAGCTTCAAGGAAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2166
D 2122 GTAGAGCTTCAAGGAAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2181
Q 2167 TGTACAGTACGCGGAAAGCTTAATTAATGCTCAATCAGTGTATGCAAAAGCTAACAC 2226
D 2182 TGCATATATTTGAAAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2241
Q 2227 AAGATTAATCTGGAAGAGTATGAGATGCTTAAGAAAGAACTGTGAGAAATTAATG 2286
D 2242 GAAAGTAAATCTAATAAGAGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2301
Q 2287 AAGAGTGAAGAGAGTGAAGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2346
D 2302 GAAAGTGAAGAGAGTGAAGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2361
Q 2347 CTTAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2406
D 2362 CTTAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2421
Q 2407 AAGTCACTTGTGTTTGTCACTATTAAGAAAACGAAAGTATGATCAAAAATGAT 2466
D 2422 AATTCATATCTGTTTATCATTCATTCATTAAGAAAGATTAATTAATTAATTAATTAATTAAT 2481
Q 2467 AGCAAAAACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2526
D 2482 AGCAAAAACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2541
Q 2527 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2586
D 2542 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2601
Q 2587 GATTTGCGACAGAGATTTTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2646
D 2602 GATTTGCGACAGAGATTTTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2661
Q 2647 GAATCAGATTCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2706
D 2662 GAATCAGATTCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2721
Q 2707 AATTAAGCTTGTGCAATCTGAAGCTCTGAGAGTGAAGCGCAGAAAGCTGACAGAA 2766
D 2722 CAGAGATTAATTTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2781
Q 2767 AGTACACAGACAACTACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2826
D 2782 AGCACAACAGACAACTACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2841
Q 2827 AATCTTTACAGACAACTACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2886
D 2842 AATCTTTACAGACAACTACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 2901

QY	ACATTCACATCTACCATCAGATCAAAATAATCATTTGACATCAAGAGCGATGAAACA	2946
Db	2902 ACCATTCACATCTACCATCAGATCAAAATAATCACTCACATCAACAGCGTTGCAACCA	2961
QY	2947 ACCAAGTGTACGACAGAGAGAAAGATGATCAGAGACGTGAACCGCAGTGAAGGGCTG	3006
Db	2962 ACCAAGTGTACGACAGG-----GGATGATTCAGACGACGTGAAGCGCAGTGAAGGATTG	3015
QY	3007 AGGATGATGGGTGCAATGTGATGAGGGGGGTGATAGTACCAATGCTTATTCCTCATG	3066
Db	3016 AAGATGATGGGTGAACCGTATGATGAGGGGGGTGATAGTACCAATGCTTATTCCTCATG	3075
QY	3067 ATTTAG 3072	
Db	3076 ATTAG 3081	

```

RESULT 6
US-10-654-416-1
; Sequence 1, Application US/10654416
; Publication No. US20040043409A1
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; TITLE OF INVENTION: glycoprotein (M5G) gene of human Pneumocystis carinii
; FILE REFERENCE: 4339-58054
; CURRENT APPLICATION NUMBER: US/10/654,416
; PRIORITY FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/09/762,724
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3042
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3042)
; US-10-654-416-1

```

Query Match	35.8%	Score 1098.6;	DB 19;	Length 3042;
Best Local Similarity	63.0%	Pred. No. 4.5e-207;		
Matches 1954; Conservative	0	Mismatches 1054;	Indels 93;	Gaps 13;

QY 2 TGGCGCGGGCGGTCAAGCGCGCAGGGCAGCGGGACACAGAAATGACATTGATGAGAGAA --- C 58

Db 2 TGGCGCGGGCGGTAAAGCGCGAGTAAACGAGACATCAGGATGATGAGAGAAAGTGC 61

QY 59 ATGTTTATGCTTAAATTCCTAAAGAGATGACTAAGTACGAGATGCCAAAAAAAC 118

Db 62 GCTTTTGGCTTAAATCTAAAGAGATTTCTAAGATGATTAATAAATGCAAGAAAAAT 121

QY 119 TAAAAAATTTGCCAGCAATTGACTGAAGCAAACTAAATATGAAACAAGTACACAGA 178

Db 122 TAGAAAAACATTGCAAAGAAATTAGGAGAACAAATTAACTCCAGAACAAAGTACATGAAA 181

QY 179 AACTTAAAGTTTGGCGAAGATGGAAGACAGATACAAATATGCAGAAAGAACTGAAGCCA 238

Db 182 ACTTAAAGATTTCTGTGATAGCAAAAAACGTATTAATAATGTAAAGAACTTAAAAA 241

QY 239 ATATTGAGAAAAATGTACTACATCAAGGAAAACTTAAGAGAACAAATTTAAAAA 298

Db 242 ATGTTGAAAAAAATGCGGTGATTTTAAAAACAGAAATTAGAGAAATTTGGTAAAAAGAG 301

QY 299 TTCAATTATTAACGATTAAGATTCGAAGAGAAATGAACAACAATGCCATTTTGTGAAG 358

Db 302 CTTCAAATTTGAAAAATGATGAGTGTACAAAAAATGAACAACAGTCTGTTTGTGAAG 361

OY	359	AGAGTGTGTTCAAAAGAACCTTAAGAATGATGGCAATCTTTGGAAATTAATGTCATCAAA	418
Db	362	AAGCATGCTC--TGAATCTTACAAAGAAATTCGACATTTAAAGAAACAAATGTTATCAGA	418
OY	419	AGAAACGTGATTAAGTTGGGAGAGAGTTCTTTTAAGACACT--TCGTAGCGATCTTA	475
Db	419	ATTAACGTGATTAAGTTAGCAAGAAAGTTCTTTTAAGAAATTAATAAGAAAGAAATTTTA	478
OY	476	ATGATCACTCATATGTGAAAAAACTTAAGAGATTGGCCCTGTCAATGGGAGGAAA	535
Db	479	AAGATTAATAATTCATGTGAAAAATAACGTGAAGTACTGTCMAAGATTAATGTCMAATGA	538
OY	536	GTGATAGCTTAACAAACTTGTGTGGAACAGAGAAAGACATGTAAGAAATTTTAATTG	595
Db	539	GTGACGAATGTGAAATTAATGTGTTGATCAAAAAAATACGTGTGATTAATCTTGTAAG	598
OY	596	AAAAAGATTAAGAGTG-----CGGTACTCTTTAAACAGATGTTTCAGCAGACTAG	646
Db	599	AAACCCACAAAGATGTGAATCTTTCAAAAAATCTTBAACGGAAATTAACAATBAAG	658
OY	647	GAAGTTTAAAAAAGAAACATGTCTTGAATTACTCGAACATGCTATTTTACATTGGAA	706
Db	659	AAGATGAACAACATAAAAAAATAATGCCCATTTATATAGAAGATCATTTTATATATG	718
OY	707	ATTTGGGAGACGACGATATTAATTAATGTAATTAATGGAGAGAAATGCCAAGAACAA	766
Db	719	AAAGTTGTGAAACCAATTCACGTGAAGTGAAGTAATGAAAAAATGTCAGAGAGAAA	778
OY	767	ATATTGCTTAATGTGCACACAGAGACCCGATTTTGTATCCACTAGCCAGAGGCTTACAATAG	826
Db	779	ATATTACTTACCATTTATTCATATTCAGGGTTTGATCTTATAGAACAGAAATTAACATTAAG	838
OY	827	CAGAGATATATAGGGCTGAGAGAGTTTATTAAGAGGTAGAGAGATGAGATTTTAATTG	886
Db	839	CAGAGAGATGACTTAGAAGGAATTTATAGAAGCAGACAGAAAGAACTCTTGTTG	898
OY	887	GAAAGATTCATTAAGAGATGCGACAGCTTTGTGGCATTTGTGATCCAAAGATTTCTAGTC	946
Db	899	GGAAACCTTTTACACACAGATGCTACTGCTTTGGTGGCAATTTTGAATTCAGATTCATCTC	958
OY	947	TTTAAAA---AAAAAGCGCAAGAAGAAATGCGAAGACCCCTTCAAAAAAGTCGAAA	1003
Db	959	TTTCACTCAAGAACTTAACAAAGAAATGTAATAAAATCTTGAAGATTAATGTAAAA	1018
OY	1004	ATCTCATGAAACATGAGGCTTTAGAAAGTTTATGTAAGAAAAATGCTTAATGATATGATG	1063
Db	1019	ATTTAAAGAAATGATATTAATTAATAAGCTCTATGCCAGATTATTAATGCAATTAAGATA	1078
OY	1064	GAACGAAAAAATGTGAGAGATTGCAAAATGATTTTAACAAAACCTTGCAAAATTTTCACTT	1123
Db	1079	AGGACAAAAAATGCCAAGAACTTAGTACGATTAATGAGAAACATGTAAATTTTTCACTT	1138
OY	1124	CAAAAGTCACTAATAATGCTCTTTTGTATCCAAAGAAAGAAATTAATGAATTTGGAT	1183
Db	1139	CAAAAACCTTAATGATTCATTTTTTTTGGCATGGAATTAATAATGATGAAATTAATAAT	1198
OY	1184	GGGAAGGGTTGCCAATTTCTTTAGCAAGAAATGTGCGAAATTTGAGTCCATATTTGTT	1243
Db	1189	GGGGGAATTTATCAACGTTTCTTAAGCAATTAAGATTTGTACAAAATTAAGATTCGATTTGTC	1258
OY	1244	TCTATTTTGAATAAAATGTCAGATGAGAAATGTCATGTAAAAATATAAGACCAAT	1303
Db	1259	TTTATTTTGAATAAAAGCTGTAG---AGGAAACTGCATGCAAGAAATATCAGAGACACAT	1315
OY	1304	GTTACAAAAGAGACTTGATGCACGGGCAATTAAGTGTCTGCAAGAAAATATGCGAGGA	1363
Db	1316	GCTACAAAGAGAGACTTGCACATTAAGCAAAATGAAATTAACAAAAAGAAATCGAGGAA	1375
OY	1364	TGTTACATGGTTCAAACAAAGCTGGCTGAAAGTTTCAACAAGATTAATGTAAGATAT	1423
Db	1376	TGCTGATGTTCCAAATTAACATGGCTTAATGTGTTTCCAAAAAAAATCATATGAATGT	1435
OY	1424	GTGAAGAACTGAAAAAGAAAACAAAGAAATTTCTCAACGATGAATTAATTTCTGT	1483

```
Db      1436 GCAAAAAAGTGAAGAAAGATTAAGAGATTTTCCGAGTAATGATTTATTTGCTTAT 1495
Qy      1484 GTGTACAGCCAGCAAAAGCAGCCGGTTCCTTACACATGATCTTCGAAATGAAAACTATCT 1543
Db      1496 GTGTACACCAATCAAAAAGCAGCTGATTTGCTTGCGATGATCTTCGATGAAAACTATCT 1555
Qy      1544 TTTTACGACAAACATGATCAAAAAGCAGATTTCCGACAGATTAATAATTTGCAAGAAAT 1603
Db      1556 TTTTGCAGACGATTTTGAACGAAAGCAGATTTCCAGTGAAAGAAAGACCTGCGAAGAT 1615
Qy      1604 TGGGAGAAAAGTCCAGATTTTAGAGAGATTTCAAAAAGAAATTAACATGGCCATGTCATA 1663
Db      1616 TATTAAAGAAATGTGAGGCTCTAAGAAAAGATTTCAAAAAAATTGAATGGCAGATGTCATA 1675
Qy      1664 CACTGAGCAGCAATGCAATGCTTGGGAGCTACAGAAATTTTAAACAGGTTTATTTG 1723
Db      1676 CATTAAGCCCAAAATTTGATCAATTTGAGAAACGCTAAAGAAATGAAAGAACTTTTACTAA 1735
Qy      1724 ATGAACAAGATACCTTTGAAAGACCAAGAAAGTTGTGTAATTAACCTTAAGAAAGAT 1783
Db      1736 ATGAACCTAAGATATTTGAAATAATCAAGAAATTTGTGAATGATTTTGAAGAGAAAT 1795
Qy      1784 GTAAATAATGCTTAGAAGAGAGATGACCGTTTCTTTTGTATGTGTCTTCCAAACG 1843
Db      1796 GCAATGAATGCTTAGAAGAGAAATGAAAGCTTTCTCTTTTATGTGCTTTGCAAAATA 1855
Qy      1844 CTACGTGTAGCTGATGTGTAAAGACGTGAAGACAGGTGTGAAGTATCAAAAAAATA 1903
Db      1856 GGCCTTGCAAAATTAATGTGTAAGATGTGAAATAATCAATGCAAAATTTTGAAAAAACA 1915
Qy      1904 TAAAGCTTCAATATATTGAATTTCTGAAATAATTAACAATAACAACACTGG 1963
Db      1916 TTAAAAAATCCAGAGATTTG-----ATAGTAATCTAAATTAAGAAATTAAG 1963
Qy      1964 AAGAAATTTGCTCTTGGCATACGATGCAATGCAATGTTTCACTTAATTTGTCAGAGTC 2023
Db      1964 GGCATATTTGCTTATTTGGACCCACACGTCATGATTTTGGACCAATTTGCCGAGATC 2023
Qy      2024 TTGAAAAA-----AATTAATGTGAAACCTTTGAAAAATTTGCAAACTTATTAAGCAAA 2080
Db      2084 AGGCTTGAAGATGCTCTCAAGTAGAGCTTCAAGGAAATTTGACTGATTAATCTAAT 2143
Qy      2081 GAGACCTTGAATATGACCTTATATTTAGTTTGAAGACATCTTGATTAAGAAAAAACT 2140
Db      2144 GTGAACTGCAATGAAAGATATTGTACAGTAGCGGAAACGTAAATTAATGCGTCAATCA 2203
Qy      2141 GCAAAACAAATCTTGATTAAGTACGTACATATGGGATCAAAACAGAAATTAACACATTA 2200
Db      2204 GTGGCTTATGCAAACTTAACACAGATTAATCTTGAAAGATGAGAGCTTGA 2263
Qy      2201 AAGTTTTTGTGTA-----ACAGTTCTACTGATTAACAAGAAACATTTTAGAG 2245
Db      2264 AGGAACCTGTGAGAAATTAAGTGAAGAGTGAAGAAAGCTGAAGCACTTAACAAG 2245
Qy      2246 ATAAACTTTGGAAAACTAGTTCAGCGTGAAGAAAAATGCAAGATTAACAAG 2305
Db      2324 AATTAGACAACCGCAGCTGATCTAAAAAAAGATTAAGACATATGAGAACTTAAGA 2383
Qy      2306 AACTTGAAGAAAGCAAAATGATTTAGAGAAAAACATTAAGAAATTAAGAAAAATAAAA 2365
Db      2384 AACGTGACAGAGAAAGCAATGAACAAGTCCAGTCTTTTGTCTCATCTATTAAAGAAAAAG 2443
Qy      2366 AGGATTAACAAAAATGCAATGAGAAACAAATTCGTTTTTCAACAATTAACAACAG 2425
Db      2444 AAGTAAATGATTAAGAAAGTAAAGCAAAACAGATTAAGAAAGCGTTTCAACGAG 2503
Qy      2426 ATATTAACCAAGAAAGG-----AGTCAACCTTAAGTACGCTTGTAG 2470
Db      2504 TTCAAGATACCAAAACATGTGAAATACTTACGAGAGGAGTTAAGAGTATCTCGTAA 2563
```

```
Db      2471 TTCAAGATATTTGACATTTT-----AACTGTAAAAAGAAATGAAAAAGTTCAAGTGA 2524
Qy      2564 CAGAAATTAGAAAGCTAAAGCAATTTGATTTGACAGAGAAAGTATTTGAAATATGATGATT 2623
Db      2525 CAGAAAAAGAAAGCAAAAGCTTTGATTTGATTTGATGACATGACATTCAGCTTTATGTAGAT 2564
Qy      2624 TGAAGGAAGATGTAATTAATTTGAAATCAGATTGCAAGAAATTAAGAGATTTGCAAGACT 2683
Db      2585 TAAAGAAACGTGTACACATCTAAAGATGATTTGCCAATTTAGAAAAAGATGTAAA---T 2641
Qy      2684 TAGAAGAAATATGCAAAAAATTAATTAAGCTTTGCGCAATCTGAAGCTCTGAGGTGA 2743
Db      2642 GTAAAGACCAAGTCAAAAGATGAAATAATATGTTTAAATAAGAACCACTGAAGATTA 2701
Qy      2744 AGCGGACAGAAACGTGACAGAAAGTACAAAGTACAACTACAAACAACAACAGCTTG 2803
Db      2702 AGCCACATGAATTAATAAAGCTTAACGAAACCAACATTAACAAGTACAGAAAGATCA 2761
Qy      2804 CCGATCCGAAGGCAACG-----AATGCAAAATCTTTACAGACAACAGACA 2848
Db      2762 AAGAGCAAGAAAAAAGCTAGAGAGACGAGAGAAATGCAAAATCTCTCAGCACAACAGACA 2821
Qy      2849 CATGGTTTACACAGACATGACACACACAGACAGTCTACTATCAATCTTACCATCAT 2908
Db      2822 CGTGGGTCACAAAAGACGTCAACCCATACAGACCTTCAAGACTTACCTCAAGATTACGT 2881
Qy      2909 CAAAAATTAATTAATGACATCAACAGAGGAGTGAACCAACCAAGTGTACAGACAGAGAG 2968
Db      2882 CAAAGATTAACATGACCTTCAAGAGGCGGTGTAAAGCTTACAAAGTGTACAGAGAGAG 2941
Qy      2969 AAGATGATGACAGAGACGTGAAACCGAGTGAAGGCTGAGAGTGAAGTGAATGTGA 3028
Db      2942 AAGATTAAGACAGAGAGAGTGAAGCCGAGTGAAGGAGTGAAGATGAGTGGAGTGTGA 3001
Qy      3029 TGAAGGGGGGTGATGATGACAAATGTTATTTTCATGATT 3069
Db      3002 TGAAGGGGGGTGATTAAGCAATGATGATTCATTCATGATT 3042
```

```
RESULT 7
US-10-654-416-3
; Sequence 3, Application US/10654416
; Publication No. US20040043409A1
; GENERAL INFORMATION:
; Applicant: Kovacs, et al.
; TITLE OR INVENTION: Identification of a region of the major surface
; TITLE OR INVENTION: glycoprotein (MS) gene of human Pneumocystis carinii
; FILE REFERENCE: 4239-58054
; CURRENT FILING DATE: 2003-09-02
; PRIOR FILING DATE: US/09/762,724
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patencin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3006
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (1)..(3006)
US-10-654-416-3
```

```
Query Match      30.2%; Score 927.4; DB 19; Length 3006;
Best Local Similarity 60.3%; Pred. No. 3,5e-173;
Matches 1859; Conservative 0; Mismatches 1131; Indels 93; Gaps 16;

Qy      2 TGGCGCGGGCGGTCAAGCGGACGAGCAGGACACAGAAATGACATTTGATGAGAACATG 61
```


QY 2213 GCAAGCTAACCCAGGATTAAGCTGGAAGAGTATGAGATGCTAGAAAGAACTCT 2272
 DB 2168 -----GCAAAAGGAAAGAAACAC---CAAGATGACATTAAGATTAGAGAAACTCT 2218
 QY 2273 GTGGAATTAATGTAAGAAAGTGAAGAAAGCTGCAAGCTTACCAAGAAATTAGAC 2332
 DB 2219 GTGAAAGCTTAATTAAGCTTAAGAAAGAAATGCTCAAAATTGAGAGACGAGCTTGAG 2278
 QY 2333 AACGGAGAGCTGATCTAATAAAGATTAATAGACATTAAGAGAACTTAAGAAAGCTGAG 2392
 DB 2279 AAGTAAAGAGGCTTGAAGAAAGAAAGAAAGTATTAAGAAAGTAAAGAGAGAG 2338
 QY 2393 AGGAAGCAATGAACAGATCCAGCTTGTGTTGTCATCTATTAGAAAGAAAGCAAGTATG 2452
 DB 2339 AAAAAGCATTGGAAGATGCAAACTTATTTATGAGAGGAAAGAGACCTGATATAT- 2397
 QY 2453 TATCAAAAGTAATAGCAAAACAGATTAAGAAATGCCGTTTCAACGGACTTCAAGATA 2512
 DB 2398 -----AATTAATTAAGTCAATTAATTAAGACTCATCTGATACACTTAAGAGAAAGGCA 2452
 QY 2513 CCACAAAACATGTGTAATAATCTAGAGAGGAGTAAAGATTAAGATTAAGCAAGATTAG 2572
 DB 2453 CAACAGAGATTT-----AACTTGTAGAGAAAGAAATGCAAAAGTCAATGTAACAGAAAG 2506
 QY 2573 AAGCTAAGCATTTGATTTGGCAGCAGAGATTTGGAAAGATATGATTTGAAGGAA 2632
 DB 2507 AATTAGCAGCATTTGATTTGGTACAGAGATTTGATCTCTATCTAGAAATTGAAGAA 2566
 QY 2633 GATGTAATTAATTTGAAATCAGATTTGCAAAATTAAGAGATTTGCAAAAGCTTGAAGAG 2692
 DB 2567 TATGTAATTAATTTGCACTGAAAGATTTGGGTTCAAAAGAAAGTGT---TGACTGTAGAGATC 2623
 QY 2693 TATGCAAAAGATTAATTAAGCTTTGCGCATCTGAAGCTTGAAGTGAAGCCGACG 2752
 DB 2624 CATGTAAGAAAGATTAAGAGAAATGTTCAACATTAAGCCACTAAAGTGAAGCAACG 2663
 QY 2753 AAACAGTGACAGAAAGTACCAAGCACTAACAACA-----ACAACAACCGTTGCG 2806
 DB 2684 AAATAGTAACTTAATAAAGCTTAACAATCAACAACAACAACAACAACAACAACAACA 2743
 QY 2807 ATCCGAAGGCAAGCAAGTGAATCTTACAGACAACAACAACAACAACAACAACAACA 2866
 DB 2744 ACGCAAGGCAAGCAAGCTGCACTCTTACAGACAACAACAACAACAACAACAACAACA 2803
 QY 2867 CGACACACACAGACGCTCTATCACTATCACTATCACTATCACTATCACTATCACTAT 2926
 DB 2804 CGACCCATTAAGCAACATCCACCAACATCTTACAGCTCAAGATTAAGCTTGACCT 2863
 QY 2927 CAACGAGGCGATGCAAAACCAACAGTGAAGCAAGAGAGAAAGATGATGACAGAGAG 2986
 DB 2864 CGACAAAGCGGTGAAGCTTACGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 2923
 QY 2987 TGAACCGAGTGAAGGCTGAGATGAGTGGGTGAAGTGAAGTGAAGTGAAGTGAAG 3046
 DB 2924 TGAACCGAGTGAAGGCTGAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2983
 QY 3047 CAATGCTATTTCTGCTCATGATT 3069
 DB 2984 CAATGACGATTTCTCATGATT 3006

RESULT 8
 US-10-654-416-15
 ; Sequence 15, Application US/10654416
 ; Publication No. US20040043409A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kovacs, et al.
 ; TITLE OF INVENTION: Identification of a region of the major surface
 ; TITLE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii
 ; FILE REFERENCE: 4239-58034
 ; CURRENT APPLICATION NUMBER: US/10/654,416
 ; CURRENT FILING DATE: 2003-09-02
 ; PRIOR APPLICATION NUMBER: US/09/762,724

; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: PCT/US99/18750
 ; PRIOR FILING DATE: 1999-08-17
 ; PRIOR APPLICATION NUMBER: US 60/096,805
 ; PRIOR FILING DATE: 1998-08-17
 ; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15
 ; LENGTH: 249
 ; TYPE: DNA
 ; ORGANISM: Pneumocystis carinii sp. f. hominis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(249)
 US-10-654-416-15

Query Match 6.0%; Score 185; DB 19; Length 249;
 Best Local Similarity 83.9%; Pred. No. 1.2e-26;
 Matches 209; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 2821 GAATGCAATCTTACAGACACAGACATGGGTTACAGACATGACACACAGC 2880
 DB 1 GAGTGCCAAATCTTGCAGACAGACAGACAGCTGGGTCAAAAGCTGACCCATATCTAGC 60
 QY 2881 ACGTCTACTATCACTATCACTATCACTATCACTATCACTATCACTATCACTATCACT 2940
 DB 61 ACTCTAGACATAGTCCACAGTCACTATCACTATCACTATCACTATCACTATCACTATCACT 120
 QY 2941 AAACCAACAGTGTACAGACAGAGAGAGATATGACAGAGACGTAAGCCGAGTGAAG 3000
 DB 121 AAGCTTACGAAGTGTACAGACAGAGAGAGATATGACAGAGAGTGAAGCCGAGTGAAG 180
 QY 3001 GGGCTGAGATGATGGGTGGAATGTATGAGAGGGGGTGAATGATGAGCAATGTTATTTCG 3060
 DB 181 GGGCTGAGATGATGGGTGGAATGTATGAGAGGGGGTGAATGATGAGCAATGATGTTTCA 240
 QY 3061 TTCATGATT 3069
 DB 241 TTCATGATT 249

RESULT 9
 US-10-668-749A-1
 ; Sequence 1, Application US/10668749A
 ; Publication No. US20040110205A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Agilent Technologies
 ; TITLE OF INVENTION: Methods and Systems for Nanopore Data Analysis
 ; FILE REFERENCE: 50112-1580
 ; CURRENT APPLICATION NUMBER: US/10/668,749A
 ; CURRENT FILING DATE: 2003-09-23
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 1
 ; LENGTH: 1300
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: synthetic construct
 US-10-668-749A-1

Query Match 3.0%; Score 91.4; DB 20; Length 1300;
 Best Local Similarity 42.0%; Pred. No. 8.4e-08;
 Matches 536; Conservative 0; Mismatches 741; Indels 0; Gaps 0;

QY 78 AAAGAAAGTGAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 137
 DB 1 AAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 60
 QY 138 ATTGATGAGCAAACTTAATTAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 197
 DB 61 AAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 120


```
QY 198 AGATGAAAAGACATACAAAATGCAAGAACTGAAGCCATATTGGAAGAAATGTC 257
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 180
QY 258 TACATCAAGAGAACTTAAAGAACATTTAAAGAAATTTGATTTATACGATTA 317
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 240
QY 318 GGATTGCAAGAGATGACAACTGCTATTTTGGAGGAGTATGTTCAAGAACT 377
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 300
QY 378 TAAAGATATTGCATCTTGAAGAAATAGTGCTATCAAAAGAACTGATTAAGTTGC 437
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 360
QY 438 GGAAGAAATTCTTTAAGAGCACTTGAGCGATCTTAATGATCAGTCATATGAAA 497
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 420
QY 498 AAAAATTAAAGAGATTTGCCCTGTCATGGGAGGAGAAAGTGAAGTTACAACTGTG 557
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 480
QY 558 TCTGAACCAAGAGACATGTAAAGATATTTTAATGAAAGAGTAAGAGTGGCTAC 617
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 540
QY 618 TCTTAAACAGATGTTTCAGCAGCACTAGAAAGTTTAAAGAAAGACATGCTTGAAT 677
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 600
QY 678 ACTGCAACATGCTATTTTACATTGGAATTCGAGAGACGATATATTAATGTAAT 737
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 660
QY 738 TGAATGGAGAGAAATGCAAGAACAAATATTTGCTTATATGCAACGAGACCGATTT 797
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 720
QY 798 TGATCCACTAGGCGAGAGCTACATAGCAGAGATATAGGCGCTGAAGATTTTATA 857
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 780
QY 858 GAGGTAGAGAGATGAGATTTTATTTGAAGATCATCTAAGATGCGACGCTTT 917
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 840
QY 918 GTTGCAATTGTCAGATCTTAGTCTTAAAGAAAGACGACAAAGAAATGCGA 977
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 900
QY 978 AGAAGCCCTTCAAAAGCTGCAAAATCTCATGACATGAGGCTTTAGAAAGTTTATG 1037
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 960
QY 1038 TAAAGAAATGCTTTAAGTATGATGAAAGAAATGTGAAGATTCGAAATGATAT 1097
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1020
QY 1098 TAAAGAAATGCTTTAAGTATGATGAAAGAAATGTGAAGATTCGAAATGATAT 1157
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1080
QY 1158 AAAAGAAATTAATGAATTTGTGATGGAGAGGTTGCCAATTTCTTACAGAGAGA 1217
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1140
QY 1218 TTGTCGAAATTTGAGCTCTATTTGTTTATTTTGAAGAAATATCTCAGATGAGAAA 1277
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1200
QY 1278 TGCATGTAAATATTAAGACAACTGTTTACAAAGAGACTTGTATGACGAGCAATTA 1337
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 1201 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1260
QY 1338 AGTGTGCAAGAAATA 1354
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1261 AAAAAAAAAAAAAAAAAAAAAA 1277
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
US-10-425-115-67687/c
; Sequence 67687, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 67687
; LENGTH: 1297
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1297)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: clone ID: MFT4577_161728C.1
US-10-425-115-67687

Query Match 2.8%; Score 87; DB 21; Length 1297;
Beet Local Similarity 44.4%; Pred. No. 6.3e-07;
Matches 327; Conservative 0; Mismatches 409; Indels 0; Gaps 0;

QY 34 ACACAGATGATGATGAGAAATGTTTACCTTTAATCTTAAGAGATGACTTA 93
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1271 AAAAAAAAAAACATTAATAAAAAAAAAAGAAATGAAANAAAAAAAAAAAAAAAAA 1212
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 94 AGTGAAGCAAGATGCAAAAAAAAACTAATAAAATATTTCCAGAAATGACAGCAAAA 153
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1211 AAAAAAAAAACAAACAAAAAAAAAAAAAAAAAAAAAAAAATTAACCAAAAAAAAAA 1152
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 154 CTAATATAGAACAGTACAGAAAACTTAAGGTTTTTGCAGAGATGAAAAAGCAGAT 213
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1151 AAAAAAAAAAAAAAAAAACAGAAAAAAAAAAAAAAAAATTAAGTGAATAAATAAACAAG 1092
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 214 ACAAAATGCAAGAACTGAAAGCCATATTTGAGAAAAATGACTACAAATCAAGAAAA 273
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1091 AAAAAAAAACAAAATTAATAAACAATAAAAAAAAAAAAAAAAACAAAAGAACCAAAAAA 1032
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 274 CTTAAGAGCAATTAATAAAAAAAAAATTCAGATTTAACGATTAAGATTCGAAAGAAAT 333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1031 AAAAAAAAAACCAAAAAAAAAAAAAAAAAAAAAAAAAACCAACCAAAANNAAGAAAAA 972
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 334 GAACAAATGCTATTTTGGAGGAGTATGTTCAAGAACTTAAGATGATTGCAAT 393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 971 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAATAAATAA 912
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 394 ACTTGAATAATTAAGTCTATCAAAAGAACGTATTAAGTTGCGAGAGAAAGTTCTTTTA 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 911 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAA 852
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 454 AGACACTTGTAGCGATCTTAATGATCAGTCATATGTGAAGAAAACTTAAGAGATT 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 851 AAAAAAAAAAAAAAAAAAAAAAAAAACCAAAAAAAAAACCAAAAGAAAGAAAAA 792
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 514 TGCCCTGTGAGGAGAGAGTGAAGTAACTTGTGTGAACCAAGAAAG 573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


Db 791 AATAAANAAAAAAGAAAAAAGATACAAANAAAAAAGAAAAA 732
Qy 574 ACATGTAAGATATTTTAATGAAAAAGATAGCGGTACTCTTAAACGATGTT 633
Db 731 AAAAAAAGAAAAAAGAAAAAAGATTAAGAAAAAAGAAAAAAGAAAAA 672
Qy 634 TCAGACACACTAGGAAGTTTAAAAAAGAAACATGCTTGAATTTCTGCAACATGCTAT 693
Db 671 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 612
Qy 694 TTTTACCTGGAAATTCGGAGACGATATTAATTTAATGTTGGAATGGAGAAAA 753
Db 611 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 552
Qy 754 TGGCAAGAACAAATA 769
Db 551 AACCATATTAATAAAA 536

RESULT 11

US-10-221-714A-461/c
; Sequence 461, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221, 714A
; PRIOR FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 461
; LENGTH: 6292
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-461

Query Match 2.8%; Score 87; DB 19; Length 6292;

Best Local Similarity 44.3%; Pred. No. 1.2e-06;

Matches 487; Conservative 0; Mismatches 605; Indels 7; Gaps 3;

Qy 83 AAGATGACTAGTACGACGAGATGCAAAAAAATCTAAAAAATTTGCCAGAAATGA 142
Db 4475 AAAAAAATATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATATA 4416
Qy 143 CTGAGAGAAATCTAAATATGACACAGTACAGAAAACTTAAAGTTTTTGCAGAAATG 202
Db 4415 AACAAAAAATTTAAAAAATTTCAAAAAAATTAATTAATTAATTAATTAATTA 4356
Qy 203 GAAAAAGAGATACAAATGCAAAAGCAATGAAAGCAATATTTGAGAAAAATGTACTCAA 262
Db 4355 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4296
Qy 263 TCAAGAGAAATCTTAAGAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 322
Db 4295 ACAAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4236

Qy 323 GCAAGAGAAATGACAAATGCTATTTTGGAGGAGATATGTTCAAAAGACTTAAAG 382
Db 4235 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4176
Qy 383 ATGATTCGAATCTTGGAAATTAAGTCTATCAAAAGAAAGTGAATGTCGAGAG 442
Db 4175 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4116
Qy 443 AGTCTCTT-TAAGAGACTCTGAGGATCTTAATGATCAATATGATGAAAAA 501
Db 4115 AACTTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4056
Qy 502 CTTAAAGAGATTTGCCCTGATGAGGAGAAAGTATGATGATTAACAACTTGCTGCTG 561
Db 4055 AACGAAACGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3996
Qy 562 AACGAGAAAGACATTAATTAATTTTAATTAATTAATTAATTAATTAATTAATTAATTA 620
Db 3995 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3936
Qy 621 TAAACAGATGTTTCAGACACATGAGAGATTTTAAAAAAGAAACATGCTTGAATTA 680
Db 3935 TAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3876
Qy 681 CGAAATGCTATTTTACATGGAATTCGAGAGACGATTAATTAATTAATTAATTAATTA 740
Db 3875 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3816
Qy 741 ATGGAGAGAAATGCAAGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 800
Db 3815 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3756
Qy 801 TCCAACTAGCCAGAGCTCAATAGCAGAGATTAATGAGCTGGAAGATTTTAAGAA 860
Db 3755 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3696
Qy 861 GGTAGAGAGATGAGATTTTATTTGGAAGAAATCATCTAAGATGCGACAGCTTTGTT 920
Db 3695 AACGACAAAAAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3636
Qy 921 GGCATTTGATCCAGA-----TTCTAGTCTTAAAAAAGACGACAAAGAGAAATGC 975
Db 3635 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3576
Qy 976 GAAAGAGCCCTTCAAAAAAGCTGCAAAATCCATGATGAATGAGGCTTTAGAAATTTA 1035
Db 3575 CAACAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3516
Qy 1036 TGTAAAGAAATGAGTTTAAGTAATGATGAACGAAAAATGTAAGATGCAAAATGAT 1095
Db 3515 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3456
Qy 1096 ATTAACAAAATTTGCAAAATTTTCACTTCAAAAGTCACTAATATGCTTTTGTATCA 1155
Db 3455 AACACAGATTAATAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3396
Qy 1156 ACAAAGGAAATATGAAA 1174
Db 3395 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3377

RESULT 12

US-10-425-115-120013/c
; Sequence 120013, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: Ia Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21 (53222) B

```
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 120013
; LENGTH: 1062
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1062)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_40935C.1
US-10-425-115-120013
```

```
Query Match 2.7%; Score 84.2; DB 21; Length 1062;
Best Local Similarity 45.5%; Pred. No. 2,1e-06;
Matches 296; Conservative 0; Mismatches 354; Indels 0; Gaps 0;
```

```
Oy 78 AAAGAGAGTGAAGTGAAGCAGAAATGCAAAAAAACTAAATAATTTGCCAAGA 137
Db 919 AAAAAAAAAAAAAAAAAATTAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAA 860
Oy 138 ATTGCTGAAGCAAACTTAATTTAGAACAGTACACAGAAAACTTAAGTTTGGCA 197
Db 859 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTA 800
Oy 198 AGATGGAAGAGAGATACAAAATGCAAGAACTGAAGCAATATTGAGAAAAATGTAC 257
Db 799 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAAAAAAAA 740
Oy 258 TACAATCAAGAGAACTTAAGAAAGCAATTAATAAAAAAAAAATTCAGATTACGATPA 317
Db 729 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAA 680
Oy 318 GGATTGCAAGAGATGAACCAATGCTATTTTTGGAGGAGATGTTCAAAAGAACT 377
Db 679 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 620
Oy 378 TAAAGATGATGCAATTAATCTTGAAGAAATTAAGTCTATCAAAAAGAACTGAATTAAGTTGC 437
Db 619 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAAAAAAAA 560
Oy 438 GGAAGAGCTTTCTTTAAGACACTTCGTAGCGATCTTAATGATGATCATATGTGAATA 497
Db 559 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 500
Oy 498 AAAAATTAAAGATTTGCCCTGTCTATGGGAGGAAAGTGTGATTAACAACCTTGTG 557
Db 499 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAGTAATAAAGAAATA 440
Oy 558 TCTGAACGAGAAAGACATGTAAAGATATTTTAATTGAAAAAGATAAGAGTGGCTAC 617
Db 439 AAGATCAAGAAATTAATAAAAAAAAAAAAAAAAAATTAAGATGTCAAAATAGATGAATTAAGCCG 380
Oy 618 TCTTAAACAGATGTTTCAACGACCTAGGAACTTTAAATAAAGAAACATGCTTTGAATT 677
Db 379 GAACACCACTATTAATAGAACCAAAACCAAGCAACAAAAACAAAGCCGCAATAGTAATG 320
Oy 678 ACTGAAACAATGCTATTTTACATTGGAATTCGCGAGCGACGATATTA 727
Db 319 ATGACGGAATTTTAAAGAAACTCTAATTAACTAGATTAAACGAA 270
```

```
RESULT 13
US-10-425-115-81106/c
; Sequence 81106, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
```

```
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 81106
; LENGTH: 1004
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1004)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_173986C.1
US-10-425-115-81106
```

```
Query Match 2.7%; Score 82.2; DB 21; Length 1004;
Best Local Similarity 46.4%; Pred. No. 5e-06;
Matches 267; Conservative 0; Mismatches 309; Indels 0; Gaps 0;
```

```
Oy 112 AAAAACTAAATAATTTGCCAAGATTGACTGAGCAAACTAAATATAGAACAACTA 171
Db 1004 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATA 945
Oy 172 CACAGAAACTTAAGGTTTTTGCAGAGTGAAGAGACAGATACAAAATGCAAGAACTG 231
Db 944 AAAAAAAAAAGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 885
Oy 232 AAGCCATATTGGAATAAATGTACTACATCAAGAGAACTTAAGAGCAATTTAA 291
Db 884 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAACCAAAAAAAAA 825
Oy 292 AAAAAATTCAGATTATPAACGATAGGATTGCAAGAGAAATGAACAACATGCTATT 351
Db 824 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 765
Oy 352 TTGAGGAGATGTCTCAAAAAGAACTTAAGATGATGCAATCTTGAGAAATTAAGTGC 411
Db 764 AATTAATAAAAAAAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAAAAAAAAA 705
Oy 412 TATCAAAAGAAAGTGAATAAGTTGCGGAGAGAGTTCTTTTAAGACCTTCGTAGCAT 471
Db 704 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 645
Oy 472 CTTATGATCATGATATGTGAATAAATACTTAAGAGATTTGCGCTGTATGGGAGG 531
Db 644 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 585
Oy 532 GAAAGTAGATTAACAACAATGTGTCTGAACGAGAAAGACATGTAAAGATTTTAA 591
Db 584 AAAAAAAAAAAAAAAAACAAAATTAATAAAAAAAAAAAAAAAAAAAAAAAAAAATA 525
Oy 592 ATTGAAAAAGTAAAGAGTGGGTACTTTAAACAGATGTTTCAAGCAGCTAGGAAT 651
Db 524 AAAAAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 465
Oy 652 TTTAAAAAAGAAACATGCTTAATTAATTAAGTAACTGACAA 687
Db 464 CAATAAGAACAAATTAATTAATTAAGTAAACAAATAA 429
```

```
RESULT 14
US-10-425-115-38710/c
; Sequence 38710, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
```

```
/ TITLE OF INVENTION: Plants
/ FILE REFERENCE: 38-21(53222)B
/ CURRENT APPLICATION NUMBER: US/10/425,115
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 369326
/ SEQ ID NO 38710
/ LENGTH: 921
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(921)
/ OTHER INFORMATION: unsure at all n locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: MRT4577_135307C.1
US-10-425-115-38710
```

Query Match 2.7%; Score 81.8; DB 21; Length 921;

Best Local Similarity 45.8%; Pred. No. 5.8e-06; Matches 319; Conservative 0; Mismatches 377; Indels 1; Gaps 1;

```
QY 71 TAAATCTAAAGAAAGATGAGTAAAGTGAAGCAAGTCAAGAACTTAAGGTT 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 846 TAAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAAT 787
QY 131 GCCAAGATTGACTGAGCAAACTAAATATAGAACTACAGAAACTTAAGGTT 190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 786 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 727
QY 191 TTTCGAGATGAGAAAGCAGATACAAATGCAAGAACTGAAAGCAATTTGAGAAA 250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 726 AAAAAAAGAAAAAATAATAATAATAATAATAATAATAATAATAATAATAA 668
QY 251 AATGTACTACATCAAGAAAGAACTTAAGAAAGCAATTAATAAATTCAGTTTAA 310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 667 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 608
QY 311 CGGATAGATGTTGCAAGAGATGAACAACATGCTATTTTGGAGGAGTATGTTCAA 370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 607 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 548
QY 371 AAGACTTAAGATGATGCTAATCTTGAAGAAATAGTCTATCAAAAGAAAGTGATA 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 547 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 488
QY 431 AAGTTGGGAAAGTCTTTTAAAGACACTTGGAGCATTTAATGATCAGTCAAT 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 487 ATACAAAAAAGAAAAAGAAAAAATAATAATAATAATAATAATAATAAAGT 428
QY 491 GTGAAAAAATACTAAAGAGTTGGCCCTGTCATGGGAGGAAAGTGATGATTAACA 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 427 AGAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAA 368
QY 551 ACTGTGCTGACACAGAAAGAGACATGTAAAGAAATATTTTAATTGAAGAAAGTAAAGT 610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 367 AATCAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAA 308
QY 611 GCGTACTCTTAAACAGATGTTTCAGACACATAGAGAGTTTAAAGAAAGAAATCATGTC 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 307 AGAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 248
QY 671 TTGAATTAATCTGACAAATGCTAATTTTAACTTGGAATTTGGGAGAGAGATTAATTA 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 247 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 188
QY 731 AATGATTTGAATTTGAGAGAAATGCAAGAAACAAA 767
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 187 AACAAAAATAGTAAAGAAAAAATGACAAAAATATAA 151
```

```
/ Publication No. US20040214272A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53222)B
/ CURRENT APPLICATION NUMBER: US/10/425,115
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 369326
/ SEQ ID NO 172717
/ LENGTH: 1243
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(1243)
/ OTHER INFORMATION: unsure at all n locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: MRT4577_89100C.1
US-10-425-115-172717
```

Query Match 2.6%; Score 81.4; DB 21; Length 1243;

Best Local Similarity 44.6%; Pred. No. 7.9e-06; Matches 292; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

```
QY 78 AAGGAAGATGACATTAAGTGAAGCAAGATCAAAAAAATAATAAATTTCCACAGA 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1134 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 1075
QY 138 ATTGACTGAGCAAACTAATAATAGAAAGTACAGAAAACTTAAGGTTTTCGA 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1074 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 1015
QY 198 AGATGAAAAAGCAGATCAAAATGCAAGAACTGAAGCAATTTGAGAAAAATGTAC 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1014 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 955
QY 258 TACATCAAGAAAGAACTTAAGAGCAATTAATAAATAAATTCAGTTTAAACGATTA 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 954 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 895
QY 318 GATTCGAAGAGATGAACAACATGCTATTTTGGAGGAGTATGTTCAAAAGAACT 377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 894 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 835
QY 378 TAAAGATGATTCATCTTGAAGAAATAGTCTATCAAAAGAAAGTGATAAGTTGC 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 834 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 775
QY 438 GGAAGAAATCTTTTAAAGCACTTGTAGCGATCTTAATGATCAGTCAATGTGAAA 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 774 AAAAAAAGTAAAAATTAATAATAATAATAATAATAATAATAATAATAA 715
QY 498 AAAAATTAAGAGATTTGGCCCTGTCATGGGAGGAAAGTGATGATTAACAATCTGTG 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 714 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 655
QY 558 TCTGACACAGAAAGAGACATGTAAAGTATTTTAATTGAAGAAAGTAAAGTCCGTAC 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 595
QY 618 TCTTAATAAGATGTTTCACAGACATAGAGAGTTTAAAGAAAGAAATCATGCTGAATT 677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 AGGAATTAATAATAATAATAATAATAATAATAATAATAATAATAA 535
QY 678 ACTGACCAATGCTATTTTAACTTGGAATTTGGGAGAGAGATTAATTAATTA 732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 ATAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 480
```

Search completed: October 29, 2005, 09:48:37
Job time : 2347 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: October 29, 2005, 06:20:41 ; Search time 10576 Seconds
(without alignments)
4686.998 Million cell updates/sec

Title: US-10-654-416-14
Perfect score: 5411
Sequence: 1 MARAYVRQAAAGTQNSIDEH.....MSGWVRGVIVAMVIFMI 1023

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=x1h
-Q=/cgn2_1/USPRO.spool/US10654416/runat_26102005_101022_21731/app_query.fasta_1.1159
-DB=GenEmbl -QMT=faetap -SUFPIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPM=pcp -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USR=US10654416.QCGN_1_17019_@runat_26102005_101022_21731 -NCPU=6 -ICPU=3
-DEV MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_da.*
2: gb_hlg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_rts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5411	100.0	3072	6	AR438440 Sequence
2	5406	99.9	3071	8	AF033211 Pneumocys
3	4398.5	81.3	3054	6	AR438439 Sequence
4	4398.5	81.3	3056	8	AF033210 Pneumocys

5	4352	80.4	3084	6	AR438437	AR438437 Sequence
6	4347	80.3	3083	8	AF033209	AF033209 Pneumocys
7	3827	70.7	3090	6	AR438436	AR438436 Sequence
8	3822	70.6	3089	8	AF033208	AF033208 Pneumocys
9	3631.5	67.1	3363	8	PMCSGI	L27092 Pneumocysti
10	3401.5	62.9	3081	6	AR438438	AR438438 Sequence
11	3396.5	62.8	3080	6	AF033212	AF033212 Pneumocys
12	2762.5	51.1	3042	6	AR438434	AR438434 Sequence
13	2762.5	51.1	12792	8	AF038556	AF038556 Pneumocys
14	2404.5	44.4	3006	6	AR438435	AR438435 Sequence
15	1731.5	32.0	3563	8	PMCSGIMS	D21827 Pneumocysti
16	1707	31.5	3511	8	PCCI17D7	CR730243 P. carinii
17	1707	31.5	36317	8	PCCI1A11	CR716157 P. carinii
18	1701	31.4	37288	8	PCC22C8	D82031 Pneumocysti
19	1687.5	31.2	11363	8	PCMO3G5	AL592382 Pneumocys
20	1668	30.8	33547	8	PMCSGF	LS5906 Pneumocysti
21	1658.5	30.7	3494	8	PMCSG910	D31909 Pneumocysti
22	1648	30.5	11370	8	PCCI21H1	CR717240 P. carinii
23	1648	30.5	25824	8	PCW01B2	AL592263 Pneumocys
24	1648	30.5	32542	8	PCCI1H1B2	CR717231 P. carinii
25	1648	30.5	35756	8	PCCI1H1B2	D31911 Pneumocysti
26	1644	30.4	5701	8	PCU83J23	US3323 Pneumocysti
27	1628.5	30.1	3242	8	PMCSGI	D17438 Pneumocysti
28	1627.5	30.1	3016	8	PMCSG8A	L29497 Pneumocysti
29	1620	29.9	3520	8	PCCI18A9	CR716158 P. carinii
30	1607	29.7	21458	8	AF043102	AF043102 Pneumocys
31	1586	29.3	4002	8	AB028489	AB028489 Pneumocys
32	1582.5	29.2	3321	8	PCU1871	U19871 Ferret Pneu
33	1581	29.2	3491	8	AF164561	AF164561 Pneumocys
34	1531.5	28.3	2982	8	AF063236	AF063236 Pneumocys
35	1502.5	27.8	3419	8	AL590130	AL590130 Homo sapi
36	1459	27.0	183557	2	AF063235	AF063235 Pneumocys
37	1439.5	26.6	3287	8	AF063238	AF063238 Pneumocys
38	1370.5	25.3	3125	8	AL590130	AL590130 Homo sapi
39	1349.5	24.9	183557	2	AF149015	AF149015 Pneumocys
40	1324.5	24.5	3459	8	AF063234	AF063234 Pneumocys
41	1302	24.1	3484	8	PMCSG2	D17439 Pneumocysti
42	1155.5	21.4	2180	8	AF063237	AF063237 Pneumocys
43	1076	19.9	2528	8	PMCSGA	M8349 Pneumocysti
44	1055	19.5	2370	8	PMCSGE	L05905 Pneumocysti
45	1052	19.4	2186	8	PMCSGE	

ALIGNMENTS

RESULT 1
AR438440
LOCUS AR438440 3072 bp DNA linear PAT 20-FEB-2004
DEFINITION Sequence 13 from patent US 6664053.
ACCESSION AR438440
VERSION AR438440.1 GI:42663295
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3072)
AUTHORS Kovacs, J.A., Huang, S., Mearns, H., Fischer, S.H., Gill, V.J. and Wei, Q.
TITLE Identification of a region of the major surface glycoprotein (MSG) gene of human *Pneumocystis carinii*
Patent: US 6664053 A 13 16-DEC-2003;
JOURNAL *Location/Qualifiers*
FEATURES
source 1..3072
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 0
Score: 5411.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Length: 3072
Matches: 1023
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

```
US-10-654-416-14 (1-1023) x AR438440 (1-3072)
QY 1 MetAlaArgAlaValIysArgGlnAlaIaGlyThrGlnAAsnSerIleAspGluGlnHis 20
Db 1 ATGGCGCGGGCGGTCAACGGCGGACAGGACACAGAAATGAGATTGATGAGAGAACT 60
QY 21 ValIeuAlaIeuIleLeuIysGluAAspGlyLeuSerGluGlnGluCysIysIysIeu 40
Db 61 GTTTTACCTTAATCTTAAGAGAGATGACTTAAGTGACAGAGAAATGCAAAAAAATACTA 120
QY 41 LysIysIysIysCysGlnGluLeuThrGlnAlaIysIeuAAsnIleGluGlnValHisArgIys 60
Db 121 AAAAAATATTCACAGAAATGACTGAAACCAAACTAAATATAGAACAACTACACAGAAA 180
QY 61 LeuIysGlyPheCysGluAAspGlyLysAlaAspThrIysCysIysGluLeuIysAlaAsn 80
Db 181 CTTAAAGCTTTTGGAGAGATGGAAAGACAGATTCAAAATGCCAAAGAACTGAAAGCCAA 240
QY 81 IleGluIysIysCysThrThrIleIysGlyIysIeuIysGluAlaIleIysIysIle 100
Db 241 ATTGAGAAAAAATGTACTACATCAATCAAGAAAACTTAAAGAAAGCAATTTAAAAAAAT 300
QY 101 GlnIleIleThrAAspIysAspCysIysGluAAsnGluGlnGlnCysIeuPheLeuGluGly 120
Db 301 CAGATTATACGAGATAGGATTGCAAAAGAAATGAACAACAAATGCTATTTTGGAGGGA 360
QY 121 ValCysSerIysGluLeuIysAAspAspCysAAsnThrIeuAArgAAsnIysCysIysIys 140
Db 361 GTATGTTCAAAAGAACTTAAGATGATTGCAATCTTTGAGAAATTAATGCTATCAAAAG 420
QY 141 LysAAspIysIysValAlaGluGluValLeuIeuAArgAlaIeuAArgSerAAspIeuAAsnGly 160
Db 421 AAACGTGATTAAGTTGCCGGAAGAGTTCTTTTAAGAGCACTTCGTAAGCATTTAATGA 480
QY 161 SerValIleCysGluIysIysIeuIysGluIleCysProValMetGlyArgGluSerAsp 180
Db 481 TCAGTCATATGTAAGAAAAAACTTAAGAGATTGCTGTCATGAGGAGGAGAAAGTAT 540
QY 181 GluIeuThrAAsnLeuCysIeuAAsnGlnIysGluThrCysIysAAsnIleIeuIleGluIys 200
Db 541 GAGTTTAACAACTTGTCTGTACCAAGAAAGACATTAAGAAATTTTAATTCGTAATG 600
QY 201 AspIysIysCysGlyThrLeuIysThrAAspValSerAlaIleuGlySerPheIysIys 220
Db 601 GATAAGAAAGTCCGGTACTTTAAAAACAGATGTTTCAGCAGCACTGAGAACTTTAAAAA 660
QY 221 GluThrCysLeuGluLeuIeuGluGlnCysIysIysIysIysIysIysIysIysIysIys 240
Db 661 GAAACATGCTTGAATTAATCTGAAACAAATGCTATTTTACATTGGAAATTCGAGACGAC 720
QY 241 AspIleIleIysCysIleGluIeuGlyIysCysGlnGluGlnAAsnIleAlaIysIysIys 260
Db 721 GATATTAATTAATGATTAATGATTTGGAGAGAAATTCCAAGAAACAAATTAATTCCTTAATG 780
QY 261 ProProGlyProAAspPheAAspProThrAArgProGluAlaIleThrIleAlaGluAAspIle 280
Db 781 CCACCAAGACCCGATTTGATCCAACTAGCCAGAGGCTACCAATGACAGAGATTAAGG 840
QY 281 LeuGluGluPheIysIysIysValIleGluIysAAspGlyValPheIleGlyIysAAsnIleIeu 300
Db 841 CTGGAGAGCTTTATTAAGAGATGAGAGAGATGAGATTTTATTAAGAAAGATCATCTA 900
QY 301 ArgAAspAlaThrAlaIeuLeuAlaIeuIeuIleGlnAAspSerSerLeuIysIysIysAsp 320
Db 901 AGAGATGGACAGACTTTGTTGSCATTTGTGATCCAAAGTTCTAGCTTAATAAAAAAAGAC 960
QY 321 AspIysGluIysCysGluGluAlaIeuGlnIysSerCysIysAAsnProHisGluHisGlu 340
Db 961 GACAAAGAGAAATGCGAAGAAAGCCCTTCAAAAAACCTCAAAATCCATCATGACATAG 1020
QY 341 AlaIeuGluSerLeuCysIysIysIysIysIysIysIysIysIysIysIysIysIysIys 360
Db 1021 GCTTAAGAAAGTTATATGTAAGAAAAATGCTTTAAGTATATGATGAGACGAAAAATGCA 1080
QY 361 GluIeuGlnAAspIleAAsnIysThrCysIysIlePheThrSerIysValThrAAsnAsn 380
Db 1081 GAATTTGCAAAATGATTAATTAACAAACTTGCAAAATTTTCACTTCAAAAGTCACTAAATAT 1140
QY 381 ArgIeuPheAAspProThrIysIysIysAAsnGlnIleValGlyTTPGluGlyIeuProThr 400
Db 1141 CGTCTTTTGTATCCAAAGAAAGAAATATGAAATTTGTGATGGAGGAGGTTCCCAACA 1200
QY 401 PheIeuSerAAsnIysAAspCysAlaIysIeuGluIysIysIysIysIysIysIysIysIys 420
Db 1201 TTTCTTGCAACAGAAATTTGTCGCAAAATTTGAGAGCTTATGCTTCTTATTTTGAAGAAAA 1260
QY 421 CysProAAspGlyIysAAsnAlaCysIysAAsnIleAArgAlaThrCysIysIysIysIys 440
Db 1261 TGTCAGATGAGAAATATGATGTAATAATTAAGACCAACATGTTCAAAAGAGAGACT 1320
QY 441 AspAlaArgAlaAAsnIysValIleuGlnGluAAsnMetArgIysMetLeuHisGlySerAsn 460
Db 1321 GATCCACGGGCAATTAAGTGTGCAAGAAATATGCGAGAAATGTTACATGTTCAAC 1380
QY 461 LysSerIrisLeuGluIysPheGlnGlnGluIysValIysValCysGluIysIysIysIys 480
Db 1381 AAAAGCTGCTTGAAGAACTTTCACAAAGATTAAGTAAAGTATGAGAACTGAAAAA 1440
QY 481 GluAAsnIysIysIysPheSerAAspAAspGluIeuPheIleIeuCysValGlnProAlaIys 500
Db 1441 GAAACAAAGAAATTTCTCAACGATGAATTAATTTATGTTGTTGTAACACCCAGCAAA 1500
QY 501 AlaAlaArgIeuIeuThrHisAAspIeuAArgMetIysThrIlePheIeuAArgGlnIleu 520
Db 1501 GCACCCGGTGTCTTAACATGATCTTCGAATGAAACTATCTTTTAACGCAACCACTG 1560
QY 521 AspGlnIysAAspAAspPheProThrAAspIysAAsnCysIysGluIeuGlyArgIysCysGln 540
Db 1561 GATCAAAAGCAAGATTTCCCGACAGATTAATAATTTGCAAGAAATTTGGGAGAAAGTCCAA 1620
QY 541 AspIeuGlyIysAAspSerIysGluIleThrTrpProCysHisIleThrIeuGluGlnCys 560
Db 1621 GATTTAGAGAGATTCAAAGAAATTAACATGCCATGTCATACATGAGACACCAATGC 1680
QY 561 AsnAArgIeuGlyThrThrGlnIleIeuIysGlnValIleuIeuAAspGluHisIysAAspThr 580
Db 1681 AATGCTTGGGAGCTTAAGAAATTTTAAAGCAGGTTTATGATGAACACAAAGATACT 1740
QY 581 LeuIysAAspGlnIysIysCysValIysIysIysIysIysIysIysIysIysIysIysIys 600
Db 1741 TTGAAAGACCAAGAAAGTGTGTAATAATCTTAATAAGAAAGTATTAATGCTTGA 1800
QY 601 ArgGlyAAspAAspAAspPheSerPheValCysValPheGlnAAsnAlaThrCysGluIeuMet 620
Db 1801 AGAGAGATGACCGTTTCTTTTGTATGTCTTCCAAAACGCTACGTTGAGAGCTGATG 1860
QY 621 ValIysAAspValIysAAspAArgCysGluValPheIysIysAAsnIleIysIysIysIys 640
Db 1861 GTAAAGACCGGAAAGCAGGTTGAAATTTCAAAAAAATTAATAAGCTTCAATATAT 1920
QY 641 IleGluPheIeuGlnAAsnIleThrAAsnIysIleThrThrIeuGluAArgAAsnCysProSer 660
Db 1921 ATTGAATTTCTTGAATAATTAATCAAAATTAATAACACATCGGAAAGAAATTTGCCCT 1980
QY 661 TrpHisThrIysCysAAsnAArgPheSerProAAsnCysProGlyIeuThrIysGluAAsnSer 680
Db 1981 TGGCATATGATTAATCAATAGATTTTTCACCTTAATTTGTCGAGGCTTACGAAAGAGATAGT 2040
QY 681 CysThrIysIleIysIysIysIysIysIysIysIysIysIysIysIysIysIysIysIys 700
Db 2041 TGTAACAAATTAACAAAGACATCGTAGCCGCTTCTATTAATAAGAAAGGCTTGAAGATGCT 2100
QY 701 LeuIysValIleGluIeuGlnGlyIysIysIysIysIysIysIysIysIysIysIysIys 720
Db 2101 CTCAAGATGAGCTTCAAGGAAATTAATGATTAATCTTAATGTAACCTGCATTGAAA 2160
```

```

Qy 721 ArgTTCrThrValAlaGlyAsnValAsnAsnAlaSerIleSerGlyLeuCyAlaVal 740
Db 2161 AGATATTGTCAGTACGCGGAGAACTTAATATATCGTCATCATGCTTATGCAAGCT 2220
Qy 741 AsnThrLyAspAsnSerGlyLySerSerAspGluAspAlaArgGlyGluLeuCySerGlyLys 760
Db 2221 AACCCAAAGATAACTCTGGAAAGATGATGAGATGCTAGAAAGAACTCTGTGAGAAA 2280
Qy 761 LeuValLySerGluValGluGluGlnCyAlaValAlaLeuProThrGluLeuGlnProAla 780
Db 2281 TTAGTGAAGAGTGAAGAGAACACTGCAACACATTACCAACAGAAATTAGGACACCGGCA 2340
Qy 781 AlaAspLeuLySerLyAspLyThrLyThrGlyGluLeuLySerLyAspGluAlaGluAla 800
Db 2341 GCTGATCTTAAAAAAGATTATTAAGCATATAGAGAACTTAAGAACTGCGAGAGAAAGCA 2400
Qy 801 MetAsnLySerSerLeuValLeuSerLeuIleLySerLyAsnGluSerAsnValSerLys 820
Db 2401 ATGAACAAGTCCAGCTCTTGTCTTCTCACTCATTAAGAAAAAGAAAGTAATGTATCAAA 2460
Qy 821 SerAsnSerLyAsnLyAspLyAsnAlaValSerAsnGlyLeuGlnAspThrThrLys 840
Db 2461 AGTAAATAGCAAAAACAAAGATMAAGATGCCCTTCAACGAGACTTCAAGATACCAAAA 2520
Qy 841 HisValLySerIleLeuArgArgGlyValLyAspValSerValThrGluLeuGluAlaLys 860
Db 2521 CATGTGAAAATACTACGAGGAGGAGTTAAGCATGTATCCGTAAACAGAAATTAGAACTTAA 2580
Qy 861 AlaPheAspLeuAlaAlaGluValPheGlyArgGlyValAspLeuLySerGluArgCyAsn 880
Db 2581 GCATTTGATTTGGCAGCAGAAATTTGGAAAGATATGTAAGTTTGAAGAAAGATGTAT 2640
Qy 881 LysLeuGluSerAspCyAlaGlyIleLySerGluAspCySerLyAspLeuGluGluValCyLys 900
Db 2641 AAATTTGAATCAGATTGCAAGATTAGAGAGATTGCAAAAGCTTAAGAAAGATGCAAA 2700
Qy 901 LysIleAsnLySerAlaCyAspAsnLeuLySerProLeuGluValLyProHisGluThrVal 920
Db 2701 AAGATTATTAAGGCTTGTGCGAATCTGAGGCTTGGAGGTGAAGCCGCGACAAACAGTG 2760
Qy 921 ThrGluSerThrThrThrThrThrThrThrThrThrThrValAlaAspProLysAlaThr 940
Db 2761 ACAGAAAGTACAAAGCAACACAAACAAACAAACAAACCGTTGCCGATCCGAAAGCAACG 2820
Qy 941 GluCySerLySerLeuGlnThrThrAspThrTrpValThrGlnThrSerThrHisThrSer 960
Db 2821 GAATGCAAAATCCTTACAGACAAACAGACACATGGGTTTACACAGACATCGACACACAAAGC 2880
Qy 961 ThrSerThrIleThrSerThrIleThrSerLySerIleThrLeuThrSerThrArgArgCys 980
Db 2881 ACGCTACTCTACATCATCATCATCATCAAAAATTAACATTGACATCAACAGCGCGATGC 2940
Qy 981 LysProThrLySerThrThrGlyGluGluAspAspAlaGlyAspValLysProSerGlu 1000
Db 2941 AAACCAACCAAGTATACGACAGAGAGAAATGATGACGAGACGCTGAAACGAGTGAAG 3000
Qy 1001 GlyLeuArgMetSerGlyTrpAsnValMetArgGlyValIleValAlaMetValIleSer 1020
Db 3001 GGGCTGAGAGATGATGGGTGAATGTGATGAGGGGGGTGATAGTATGCAATGTTATTTCCG 3060
Qy 1021 PheMetIle 1023
Db 3061 TTCATGAT 3069

```

```

RESULT 2
AF033211 3071 bp DNA linear PLN 10-SEP-1998
LOCUS AF033211
DEFINITION Pneumocystis carinii f. sp. hominis clone HUMSG35 major surface
ACCESSION AF033211
VERSION AF033211.1 GI:3560518
KEYWORDS

```

```

SOURCE
ORGANISM
Pneumocystis jirovecii
Pneumocystis jirovecii
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.

REFERENCE
AUTHORS
1 (bases 1 to 3071)
Garbe, T.R. and Stringer, J.R.
TITLE
Molecular characterization of clustered variants of genes encoding
major surface antigens of human Pneumocystis carinii
JOURNAL
Infect. Immun. 62 (8), 3092-3101 (1994)
MEDLINE
94314421
PUBMED
7518806
REFERENCE
AUTHORS
2 (bases 1 to 3071)
Mei, Q., Turner, R.E., Social, V., Kivington, D., Angus, C.W. and
Kovacs, J.A.
TITLE
Characterization of major surface glycoprotein genes of human
Pneumocystis carinii and high-level expression of a conserved
region
JOURNAL
Infect. Immun. 66 (9), 4268-4273 (1998)
MEDLINE
98380374
PUBMED
9712777
REFERENCE
AUTHORS
3 (bases 1 to 3071)
Mei, Q., Turner, R., Social, V., Kivington, D., Angus, C.W. and
Kovacs, J.A.
TITLE
Direct Submission
Submitted (07-NOV-1997) Critical Care Medicine Dept., National
Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD
20892-1662, USA
FEATURES
source
Location/Qualifiers
1..3071
/organism="Pneumocystis jirovecii"
/mol_type="genomic DNA"
/species="Homo sapiens"
/sub_species="hominis"
/db_xref="taxon:42068"
/clone="HUMSG35"
/note="derived from HIV-infected human with P. carinii
pneumonia"
<1..3071
/gene="MSG"
<1..3071
/gene="MSG"
/gene="MSG"
/note="surface antigen"
/codon_start=3
/product="major surface glycoprotein"
/protein_id="AAC34974.1"
/db_xref="GI:3560519"
/translacion="AAVVRQAAGTQNSIDBEHVLAALIKEDGISEQCKKLLKCYC
ELTEAKINIQVHRKLGFCDEGADADPKCELRANI EKKCTTIGKLUKEAIKKKIQII
TDKCKENEOQFLBEGVCSKELDKNTLRNKCYQKRDKVAEVLRLARSLDINS
VICERKLKEICPMWGRESDELTMNLCINQKCTKNIIIEKDKGTLTQVSAALGFEK
KENTCEILBOCYFYIGNCGDDI IKCIEIDGKQOENIAAMPPEPDPRPATIE
DIGIEEPYKVEBQVITGKMLRDATALLALIODSLKKDKKCEERLQSCSN
PHEHEALPESLCKKNGLSNDGTCKCEBLQNDINTCKIPTSRTNNRI.PDPKNNBIV
GWGGLPTFLSNEOCALKESYCFYEKKCPDGENACNRIATCYRGGIDARANKVLOEN
MRGMLHSKMSMLEKFOELVYKCEKLEKKNKGSFNSDELFLICVQPAARLITLHL
RMKTIPLRQOLDKRPDPTDKCKELRKGODIGEDSKETIWPCHTPOCCNRGTTE
ILKQVILDEBKDITLKQESCVYIKCEKCMKMSRSDRSPVCFQFQATCELMKQDK
DRCEVFKKNIKASITIEFLNNTRKITTLENNCPSWHTYCNRPSPNGILTKENSCKK
IKHREPFYRKLIEDLKVLEQKLDKSKCEPALRYCTVAGVNNASISGICRAN
TKNNGSGSDADARKEICELKVEVEBQKALPELGPADILKVDYTYBELKKRAE
ANMKSILVSLIKKESNVSNSKNSKNDRKAVNSGLDITKHVILIRGVADVSTEL
EAKAFDLAEEVFGVYDLKERCNKLSEJDKEDCQDLSEYCKINKACNKLKLEVK
PHEVTESTTTTTTTTTTVAADPKATECKSIQTDITWTQTSHTSTSTSTSTSKIT
LITSTRCKPTKCTTGGEDDAGDYKPSBGLMKSGMNVNRGIVLVNVISFMI"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 0
Score: 5406.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 99.91%
DB: 8
Length: 3071
Matches: 1022
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

```

US-10-654-416-14 (1-1023) x AF033211 (1-3071)

2 AlaArgAlaValIlysaRgInlaIaIaGlyThrGlnAenSerIleAspGluGluHleVal 21
3 GCGGGGGGGGTCAAGCGGCGAGCAGGACACAGAAATAGCATTTGAGAGAAACATTT 62
22 LeuAlaLeuIleLeuIlyGluAspGlyLeuSerGluGlnGluCysLysLysLysLys 41
63 TTAGCTTAAATTTCTTAAGAGAAATGAGCTTAAGTGACAGAAATGCAAAAAAATTANA 122
42 LysTyrCysGlnGluLeuThrGlnAlaLysLeuAenIleGluGlnValHisArgLysLeu 61
123 AAATATTCACAGAAATTAAGTGAACCAAACTAAATATATAGAACAAATTAAGCAAACTT 182
62 LysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAlaAenIle 81
183 AAAGCTTTTTCGAAAGATGCAAAAGCAGATACAAAATCCAAAAGAACTAAATTT 242
82 GluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLysLysIleGln 101
243 GAGAAAAAATGTACTACATCAATCAAGGAAAACTTAAAGAACATTAATAAAAAAATTTCAG 302
102 IleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeuPheLeuGluGlyVal 121
303 ATTAATACGGAATAGGATTTGCAGAAAGATGAACAAATGCTATTTTTCGAGGAGATA 362
122 CysSerLysGluLeuLysAspAspCysAsnThrLeuAArgAenLysCysTyrGluLysLys 141
363 TGTTCAAAAGAACTTAAAGATGATTTGCATATCTTTGAGAAATAGTGTATCAAAAAGAA 422
142 ArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGlySer 161
423 CGTGATAAAGTTGCGGAGAAAGTTCTTTAAAGACACTTCGAGGATCTTAATAGATCA 482
162 ValIleCysGluLysLysLysLysGluIleCysProValMetGlyArgGluSerAspGlu 181
483 GTCATATGTGAAAAAACTTAAGAGATTTGCCCTGTCAATGGGAGGAAAGTATAG 542
182 LeuThrAenLeuCysLeuAsnGluLysGluThrCysLysAsnIleLeuIleGluLysAsp 201
543 TTAAACAACTTGTCGTCTGACCAAGAAAGACATGTAAGATATTTTAATGAAAAAGAT 602
202 LysLysCysGlyThrLeuLysThrAspValSerAlaAlaLeuGlySerPheLysLysGlu 221
603 AAGAAAGTCGGTACTTTAAACAAATGTTTCAGACGACTAGGAAAGTTTAAAAAAANA 662
222 ThrCysLeuGluLeuLeuGluGlnCysTyrPheTyrIleGlyAenCysGlyAspAspAsp 241
663 ACATGCTTGAATTACTGAAACAAATGCTATTTTTCATTTGAAATTTGCGAGACAGCAT 722
242 IleIleLysCysIleGluLeuGlyGlyLysCysGlnGluAenIleAlaTyrMetPro 261
723 ATTAATTAATGATTAATGAAATGGAGGAAATGCCAAAGAACAAAATTAATGCTTAATGCA 782
262 ProGlyProAspPheAspProThrArgProGluAlaThrIleAlaGluAspIleGlyLeu 281
783 CCAGAGCCCGGATTTTGATCCAACTAGCCAGAGCTTACAATAGCAGAGAAATATAGGCTG 842
282 GluGluPheTyrLysLysValGluGluAspGlyValPheIleGlyLysAsnHisLeuArg 301
843 GAAAGGTTTATAGAGAGGATGAGGAGATGAGTTTATTTGGAAGAAATCATATTAAGA 902
302 AspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLysLysLysAspAsp 321
903 GATGGAGACAGCTTTGCTTGCAATTTGTGATCCAAAGTTCTAGCTTAAAAAAAACACAC 962
322 LysGluLysCysGluGluAlaLeuGlnLysSerCysLysAspProHisGluHisGluAla 341
963 AAAGGAAATGCGAAGAAAGCCCTTCAAAAAGCTCAAAAATCCCTCATGAACATAGGCT 1022
342 LeuGluSerLeuCysLysLysAsnGlyLeuSerAsnAspArgIleThrLysLysCysGluGlu 361

1023 TTAGAAAATTATGTAGAAAAAATGCTTAAAGTATATGTAAGAAAAAATGTCAGAA 1082
362 LeuGlnAsnAspIleAenLysThrCysAlaIlePheThrSerLysValThrAsnAsnArg 381
1083 TTGCAAAATGATATTAACAAAATCTTGCAAAATTTTCACTTCAAAAAGTCACTAATATATG 1142
382 LeuPheAspProThrLysGlyAsnAsnGluIleValGlyTyrGluGlyLeuProThrPhe 401
1143 CTTTATTAATCCMAAAAGAAATTAATGAATTTGTGAGTGGAGAGGTTCCCAACATTT 1202
402 LeuSerAsnGluAspCysAlaLysLeuGluIleSerTyrCysPheTyrPheGluLysLysCys 421
1203 CTTAGCAACAGAAATTTGTGCAAAATTTGAGTCTTAATTTCTTATTTTGAABAAAAAGT 1262
422 ProAspGlyGluAenAlaCysLysAsnIleArgAlaThrCysTyrLysArgGlyLeuAsp 441
1263 CCAATGAGAGAAATTAATGATGTAAATTAATAGACCAACATGTTCAAAAAGAGACTGAT 1322
442 AlaArgAlaAenLysValLeuGlnGluAsnMetLysGlyMetLeuHisGlySerAsnLys 461
1323 GCACGGGCAATTAAGTGTGCAAGAAATATGCAAGAAATGTATACATGTTCAAAACAA 1382
462 SerTyrLeuGluLysPheGlnGlnGluValLysValCysGluLysLeuLysLysGlu 481
1383 AGCTGCTTGAATTAAGTTTCAACAAAGATTAATTAATGATGAGAAATGAAAAAGAA 1442
482 AsnLysGlySerPheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLysAla 501
1443 AACAAAGAAATTTCTCAACAGATGAATTTATTTCTGTGTGTATGACCCAGCAAAAGCA 1502
502 AlaArgLeuLeuThrHisAspLeuArgMetLysThrIlePheLeuArgGlnGluLeuAsp 521
1503 GCCCGGTTGCTTACACATGATCTTCGATGAATGAACATCTTTTTCACCAACACTGAT 1562
522 GluLysArgAspPheProThrAspLysAsnCysLysGluLeuGlyArgLysCysGluAsp 541
1563 CAAAAGCAGATTTCCGACAGATMAAAATTTGCAAGAAATTTGGGAGAAATGCCCAAGAT 1622
542 LeuGlyGluAspSerLysGluIleThrTyrProCysHisThrLeuGlnGlnCysAsn 561
1623 TTAGAGAGATTCATAAAGAAATTAACATGGCCATGTCTATACATGAGACCAATGCAAT 1682
562 ArgLeuGlyThrThrGluIleLeuLysGlnValLeuLeuAspGluHisLysAspThrLeu 581
1683 CGCTTGGGAGTCAAGAAATTTTAAGCAGAGTTTATTTAGTGAACCAAAAGATACCTTGT 1742
582 LysAspGlnGluSerCysValLysTyrLeuLysGluLysCysAsnLysTyrSerArgArg 601
1743 AAAGACCAAGAAATTTGTGTAAATTAACCTAATAAAGAAATGATTAATGCTTAAGAA 1802
602 GlyAspAspArgPheSerPheValCysValPheGlnAsnAlaThrCysGluLeuMetVal 621
1803 GGAGATGACCGTTCTCTTTTGTATGTGTTCCAAAACGCTACGTTGAGACTGATGTA 1862
622 LysAspValLysAspArgCysGluValPheLysLysAsnIleLysAlaSerTyrIleIle 641
1863 AAAGACGTGAAGACAGGTGTGAAGTATCAAAAATAATATMAAAGCTTCAATATTAATT 1922
642 GluPheLeuGluAenAsnThrAsnLysIleThrThrLeuGluArgAsnCysProSerTyr 661
1923 GAATTTCTTGAATTAATTAACAAATTAATTAACAACTGGAAGAAATTTGCCCTCTTGG 1982
662 HisThrTyrCysAsnArgPheSerProAsnCysProGlyLeuThrLysGluAenSerCys 681
1983 CATAGTATTCATTAATGATTTTTCATTAATTTGTCCAGGTCTTAAGAAAGAAATGTTGT 2042
682 ThrLysIleLysLysHisArgGluProPheTyrLysArgLysAlaLeuGluAspAlaLeu 701
2043 ACAAATAATCAAGAAAGCATCGAGCCGTTCTATMAAAGAAAGCCCTTGGAAGATGCTTC 2102
702 LysValGluLeuGlnGlyLysLeuThrAspLysSerLysCysGluProAlaLeuLysArg 721
2103 AAAGTAAAGCTTCAAGAAATTAATGATTAATCTAATATGTAACCTGCATTGAAAAAG 2162

QY 722 TycSerThrValAlaGlyAsnValAsnAspAlaSerIleSerGlyLeuCyblyAsn 741
 Db 2163 TATTGTACAGTACGGGAAACGTTAAATTAATCCGCAATCAGTGGCTTATGCAAGCTAAC 2222
 QY 742 ThrlyAspAsnSerGlyLySerSerPduAspAlaArglyGlyLeuCyblyLyLeu 761
 Db 2223 ACCAAGATTAACCTCGAAGAGATGATGAGATGCTGAAAGAACTCTGAGAAATTA 2282
 QY 762 VallyGlyValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 781
 Db 2283 GTGAAAGAGTGAAGAAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGT 2342
 QY 782 AspLeuLeuValAspPyllySerThrThrGlyGlyLeuLeuValAspAlaGlyValAlaMet 801
 Db 2343 GATCTAATAAAGATTTATTAACATATGAGAACCTTAAGAACTGCAAGAGAACTAATG 2402
 QY 802 AsnlySerSerLeuValLeuSerLeuIlelyblyAsnGlySerAsnValSerlySer 821
 Db 2403 AACAAAGTCCAGTCTGTTTGTCTCACTATTAAAGAAAAAGAAAGTAAATGATCAAAAAGT 2462
 QY 822 AsnSerlyAsnlyAspPyllyAsnAlaValSerAsnGlyLeuGlnAspThrThrlyHis 841
 Db 2463 AATAGCAAAAACAAAGATTAAGAAAGCCGTTCAACGCACTTCAAGATACCAACAAACAT 2522
 QY 842 VallySerIleuAspArgGlyValIlyAspValSerValThrGlyLeuGlyAlaIlyAla 861
 Db 2523 GTGAAATTAACCTCGAAGAGGAGTTTAAGATGATATCCGTAAAGAAATTAAGAACTTAAGCA 2582
 QY 862 PheAspLeuAlaAlaGlyValIlePheGlyArgTyValAspLeuValGlyIleGlyAsnly 881
 Db 2583 TTTGATTTGGCAGCAGAAAGATTTTGAAGATATGATTTTGAAGAAAGATGTAATTA 2642
 QY 882 LeuGlySerSerPyllyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 901
 Db 2643 TTGAAATCAAGATTCAGAAATTAAGAGAGATTTGCAAGAGCTTAAGAAAGTATGCAAAAAG 2702
 QY 902 IleAsnlyValIleCyblyAsnLeuLeuProleuGlyValIlyblyProHsglyThrValThr 921
 Db 2703 ATTATATAGGCTTGTCCGCAATCTGAAGCTTCTGAGGTTGAAGCCGCAAGAAAGCTAACA 2762
 QY 922 GluSerThrThrThrThrThrThrThrThrThrThrValAlaAspProlyblyAlaThrGly 941
 Db 2763 GAAAGTCAACGACAACTCAACAACAACAACAACCGTTCGATCCGAAAGCGCAACGGA 2822
 QY 942 CyslyblySerLeuGlnThrThrAspThrTTPValThrGlnThrSerThrHsrThrSerThr 961
 Db 2823 TGCAAATCCTTACAGAACACAGACACATGGGTTACACAGACATCAGACACACAAAGCAGC 2882
 QY 962 SerThrIleThrSerThrIleThrSerlyblyIleThrLeuThrSerThrArgGlyCysly 981
 Db 2883 TCTACTATCACTTACATCACTCAAAAATTAACATTGACATTAACGAGCGATGCAAA 2942
 QY 982 ProThrlyblyCysThrThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1001
 Db 2943 CCAACCAAGTATGACAGACAGAGAGAGATGATCAAGAGACGTGAACCAAGTGAAGGCG 3002
 QY 1002 LeuArgMetSerGlyTTPAsnValMetArgGlyValIleValAlaMetValIleSerPhe 1021
 Db 3003 CTGAGGATGAGTGGTGAATGTGATGAGGCGGTGATAGTGAACCAAGTATTTCTTC 3062
 QY 1022 MetIle 1023
 Db 3063 ATGATT 3068

RESULT 3
 LOCUS AR438439 3054 bp DNA linear PAT 20-FEB-2004
 DEFINITION Sequence 11 from patent US 6664053.
 ACCESSION AR438439
 VERSION AR438439.1 GI:42663294
 KEYWORDS Unknown.

ORGANISM Unknown.
 Unclassified.
 REFERENCE 1 (bases 1 to 3054)
 AUTHORS Kovacs,U.A., Huang,S., Meaur,H., Fischer,S.H., Gill,V.J. and Mei,Q.
 TITLE Identification of a region of the major surface glycoprotein (MSG)
 JOURNAL gene of human Pneumocystis carinii
 PATENT: US 6664053-A 11 16-DEC-2003;
 FEATURES Location/Qualifiers
 source 1..3054
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
 Alignment Scores:
 Pred. No.: 2,25e-274 Length: 3054
 Score: 4398.50 Matches: 838
 Percent Similarity: 88.81% Conservative: 75
 Best Local Similarity: 81.52% Mismatches: 98
 Query Match: 81.29% Indels: 17
 DB: Gaps: 9

US-10-654-416-14 (1-1023) x AR438439 (1-3054)
 QY 2 AlaArgAlaValIlyblyArgGlnAlaAlaGlyThrGlnAsnSerIleAspGlyLeu--His 20
 Db 1 GCGGGGCGGTCAACGGCGAGTAAACAGAGCATCAGGCAATGATGATGAAGTGAAT 60
 QY 21 ValLeuAlaLeuIleLeuLyblyGlyblyLeuSerGlyLeuGlnGlyCyslyblyLeu 40
 Db 61 ATTTTGGCGTTGATTTCTACAAAGAGATGATCAATGATCAAAATGCAAAAAGTTTA 120
 QY 41 Lyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 60
 Db 121 GAAATTAATCTGCAAGAGTTGAAAAAGCACTACTAGACATGAAAAAGTATCAATAATG 180
 QY 61 LeuLyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 77
 Db 181 CTTAAAGATTTCTGTGAAATGGAATGGAAGCAAGTAAACAAATCAAAATGCTCAAGTCTA 240
 QY 78 Lyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 96
 Db 241 CAAGCCAAAGTTACGGGGAATGTACAAATTTTAAACAAAGAGTGAAGCAAGCTTA 300
 QY 97 Lyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 116
 Db 301 ACAATCCA-----TCAGATGATTAATTTGCAAGAGAGTGAACCAATAGCTTA 348
 QY 117 PheLeuGlyGlyValIlyCysSerlyblyblyblyblyblyblyblyblyblyblyblybly 136
 Db 349 TTTTGGAGGAGAGATGC---CATTAATCTTGTGAAGATTTGTAAACAATTAAGAAATCTA 405
 QY 137 CysTyGlnlyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 156
 Db 406 TGTTAACCAAGAAAAACCTGACGAGTGAAGAGTACTCTTTTGAAGGAGCACTTCGATG 465
 QY 157 AspLeuAsnGlySerValIleCysGlyblyblyblyblyblyblyblyblyblyblybly 176
 Db 466 GATCTCAATTAACAGAAACACATGAAAAAAATGAAAGAGATTTCCTCACTTTCAG 525
 QY 177 ArgGlySerAspGlyLeuThrAsnLeuCyblyblyblyblyblyblyblyblyblyblybly 196
 Db 526 AGGGAAGTATGATTAACGAGACTGTGTTGAACCAAGAAAAAGAGCTGCAAGATATT 585
 QY 197 LeuIleGlyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 216
 Db 586 ATAAAGAAAAAGATTAATAATGCACTTAAGCAATGTTGCAACCAAGCTTGA 645
 QY 217 SerPheLyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblyblybly 236
 Db 646 AGTTTAAAAAGAAATATGCTTGAATTACTTGAACAAAGCTATTTTTCATTGCAAT 705
 QY 237 CysGlyAspAspAspIleIlelyblyblyblyblyblyblyblyblyblyblyblybly 256

[illegible]

OY	616	ThrcYsgLlueWetValIlysaapValIyaaapAxcYsgLlualPheIylysaAIIle	635
Db	1834	ACGtGTAGcCTGAGTGAAGACGTGAAGACAGGtGTAGATTCACAAAAAATATA	1893
OY	636	LySaIaSerYtIleIleGluPheLeuGlaAaSnThraSnIySIIethrThLeuGlu	655
Db	1894	AAACCTTCATATATATGAAATTCCTTGAAATATATACAAATAAATACACACTGGA	1953
OY	656	ArgaSnCySProSerTTPHLeThrYTCyAaSnAxpPheSerProaSnCySProGlyLeu	675
Db	1954	AGAAATTCCTCCTTCCTGGCAACGATATGGCAATATGATTTCCACTAATTCCTCAGGCTT	2013
OY	676	ThrlrYsgLlueAaSerCySThrlrIlyIleLyLeuHIAaArgIuProPheYTLVsaYLyS	695
Db	2014	ACGAAAGGAATAGTGTATACAAAATATGAAGACATTCGTAGCCGTTCTATTAAGAAAG	2073
OY	696	AlaLeuGluAapAlaLeuLySValGluLeuGInGlyLySLeuThraPlySserLyS	715
Db	2074	GCCTTGAAGATGCTCTCAAGATAGACCTTCAAGAAAATTTGACTGATTAATGTAATGT	2133
OY	716	GluProAlaLeuLySaYgTYTCySThrValAlaGlyAaSnValaSnnaIaSerIIeSer	735
Db	2134	GAACCTGCATTTGAATAATATTGACAGTGCAGGGAACCTTAATATATGCGTCAATCAGT	2193
OY	736	GlyLeuCySlySaIlaaSnThrlrLySaapAaSerGlyLySserAxpGluAapAlaAxyLyS	755
Db	2194	GCGTTATAGCAAGCTTAACCAAGATACCTCGGAAGAGTGATGAGATGCTAGAAAG	2253
OY	756	GluLeuCySgIuLySLeuValIySgIuValGluGluGInCySlySaIaleuProthGlu	775
Db	2254	GAACCTGTGTGAAATTAAGTGAAGAAAGATGGAAGAACAGTGAAGCATTAACCAACGAA	2313
OY	776	LeuGlyGluProAlaAlaAapLeuLySlySaapTYTLySThrYrGluGluLeuLyS	795
Db	2314	TTAGACAAACCGGACGCTGATTTAAAAAGATTATAGCAATATGAGGAATTAAGAAA	2373
OY	796	ArgAlaGluGluIlaMeCaSnLySserSerLeuValLeuSerLeuIleLySlySaSnGlu	815
Db	2374	CGTCAAGAGAGCAATGAACAACTCCAGTCTTGTTGTGCATCTCAATTAAAGAAAAGAA	2433
OY	816	SerAaSnValSerIySserAaSnSerLySaSnLySaapLySaSnAlaValSerAaSnGlyLeu	835
Db	2434	AGTATATGATCAAAAAATTAATAGCAAAAAACAAGATTAAGATGCCGTTTCAAAACGGACTT	2493
OY	836	GlaaPThrThrlrYsHIsValLySIIleuAaArgGlyValIlySaapValSerValThr	855
Db	2494	CAAAATATCCAAAAACATGTGAATAATCTACGGAGAGAGAGTTAAGATGATCCGTACAA	2553
OY	856	GluLeuGluIlaLySaIaPheAapLeuAlaAlaGluValPheGlyAxyTYrValaAapLeu	875
Db	2554	GAATTAAGAGTAAAGCATTTGATTTGGCAGCAGAAATATTTGGAAAGATATGTAGATTGG	2613
OY	876	LySgIuAxpCySaSnLySLeuGluSerAapCySaYgIIeLySgIuAapCySlySaapLeu	895
Db	2614	AAGAAAGATGTATTAATTAATGGAATCGATTTGCGAATTAAGAGCGATTTCCAAAGACTTA	2673
OY	896	GluGluIlaValCySlySaIySIIeAaSnLySaIaCyAaArgaSnLeuLySProLeuGluValLyS	915
Db	2674	GAAGAAGATGCAAAAAAGATTAAATGAAGCTTGTCGCAATCGAAGCCTCGAGAGTAAAG	2733
OY	916	ProHIsGluThrValIThrGluSerThrThrThrThrThrThrThrThrValAla	935
Db	2734	CCGCAACGAACAGTGAAGAAAGTACAAACGACAACTCAACAACAACAACACGCTTGCC	2793
OY	936	AaPProLySaIaThrGluCySlySserLeuGInThrThraPThrTrValIThrGInThr	955
Db	2794	GATCCGAAGCAACGAGATGCAATCCTTACAGCAACAACAACATGCGTTTACACAGACA	2853
OY	956	SeThrHIsThraSerThrThrThrThrIleThSerThrlrIleHIsSerLySIIeThrLeuThr	975
Db	2854	TCGACACACAAACGACGCTACTATATACATCTACCAATCAATCAAAAAATTAACATTGACA	2913

QY	976	SeethiragaCyVelyperothlygcyerthrhrglygluglunapapalaglyasp	995
Db	2914	TCACGAGGGCGATGAAACCAACCAAGTGTACGACAGG-----GATGATGCAGAAAGC	2957
QY	996	VallypProsergluglyLeuArgmetSerglyTrpaenValmetAglyVallleval	1015
Db	2968	GTGAAGCCAACTGAAGCCTTAGAGGTGACCGGCTGAATGTGATGAGGGGTATAGTA	3027
QY	1016	AlameVallleSerPheMetile	1023
Db	3028	GCAATGATTATTCGTCATGATT	3051
RESULT 4	AP033210	AP033210	3056 bp DNA linear PLN 10-SEP-1998
LOCUS	DEFINITION	Pneumocystis carinii f. sp. hominis clone HUMSG33 major surface glycoprotein (MSG) gene, partial cds.	
ACCESSION	VERSION	AP033210	
KEYWORDS	ORGANISM	AP033210.1 GI:3560516	
REFERENCE	AUTHORS	Pneumocystis jirovecii Pneumocystis jirovecii Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes; Pneumocystidaceae; Pneumocystis.	
REFERENCE	TITLE	1 (bases 1 to 3056) Garbe,T.R. and Strlinger,J.R. Molecular characterization of clustered variants of genes encoding major surface antigens of human Pneumocystis carinii Infect. Immun. 62 (8), 3092-3101 (1994)	
JOURNAL	MEDLINE	94314421	
PUBLISHED	REFERENCE	7518906	
AUTHORS	TITLE	2 (bases 1 to 3056) Mei,Q., Turner,R.E., Social,V., Kilmington,D., Angus,C.W. and Kovacs,J.A. Characterization of major surface glycoprotein genes of human Pneumocystis carinii and high-level expression of a conserved region Infect. Immun. 66 (9), 4268-4273 (1998)	
JOURNAL	MEDLINE	98380374	
PUBLISHED	REFERENCE	9712777	
AUTHORS	TITLE	3 (bases 1 to 3056) Mei,Q., Turner,R., Social,V., Kilmington,D., Angus,C.W. and Kovacs,J.A. Direct Submision Submitted (07-NOV-1997) Critical Care Medicine Dept., National Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD 20892-1662, USA	
FEATURES	SOURCE	Location/Qualifiers	
		1..3056	
		/organism="Pneumocystis jirovecii"	
		/mol_type="genomic DNA"	
		/specific_host="Homo sapiens"	
		/sub_species="hominis"	
		/db_xref="taxon:42068"	
		/clone="HUMSG33"	
		/note="derived from HIV-infected human with P. carinii pneumonia"	
		<1..3056	
		/gene="MSG"	
		<1..3056	
		/gene="MSG"	
		/note="surface antigen"	
		/codon_start=3	
		/product="major surface glycoprotein"	
		/protein_id="AAC34973.1"	
		/db_xref="GI:3560517"	
		/translation="AAVARGVVTGASGYDENVILALIIQEDAMEDTKCKSLERYC BELKASLDMEKVKHMLKDFCGNGASVANTKCGQLAKVYKCTNFQKLGALNTNKG PSDNCKSEBRCQFLFGACNLVEDCKNLKCYOKRGDAEVLRLARSDLNKTKL ETHKHKKESICPVLQRESNELTDLCLNCKTCENIKKKDKCTTLKANVATLGGSP KEICBELLEQCYFYTGNGGDDDIKCTELGSKCOEQNIYIYPRGDPDPRPPTATLAE TDIDDELTKAAEEGVPTGKRLHDAITLTLVKDDDTGKNNIIGKCNKILKEDCK NSQDHEALKNLGNNSPVAYPEKCKLEBDIKTCTNLTPTLLKNHLYDNNDKIVEM	
gene			
CDS			

ORIGIN			
<p> RXKPTFLTNDCAKLRSYCYCYEACACENACBECNNLEAACYKGLDRAKVLVDENNR GLARGSQSLNKRFGQRLVYVCKEIKENKSSPNDEIFVLCYOPAKARLITDHOHR VIFLRQOLDQRPDPTDCKCEKJAKCKODJAKSKETITWCHTLEBOCNRLETTEILK QVLDERHKDITLKDOESCVKYLKEKCNMSRGRDFSVCFQVATLELWKDVKDC EVLKFNKASIAIIEFLENNTNKTITLERNPSPMHTYCNRS PNCPGITKNSCTRIKK HCEPFYRKALDELKVELOGLDKSCDPALEKRYVAGVNNASISGLCKANTDQ NSGSGEDBARKEILCEKLVKEVEBOCKALPELELOPADLKDQYVTELEKRAEANA KSSIVLSLTKCNSSNVKSNKSKYKDAVNGIOLPTTKYKTLIRGQVNVTELEAK AFSDIAVSPGRYVDLCKRKNESDCKIREDCKDQLEVCCKINKACNKLPLEVKPHE TVFESTTTTTTTTTTVADPKRATCKSKISQTTDVTQSTHSTISITSTITSKITLTS TRCKCPKCTGDDADBEVKPSEGLRVS GNMVWGVIVANVISFMI" </p>			
Alignment Scores:			
Pred. No.:	2,25e-274	Length:	3056
Score:	4398.50	Matches:	838
Percent Similarity:	88.81%	Conservative:	75
Best local Similarity:	81.52%	Mismatches:	98
Query Match:	81.29%	Indels:	17
Ds:	8	Gaps:	9
US-10-654-416-14 (1-1023) x AF033210 (1-3056)			
QY	2	AlaaRgaIaValVysArGInAlaIaGlyThrGlnaSerIleAspGluGlu--His	20
Ds	3	GCGGGGGGGGTCAAGCGGACGGACGTAAACAGACATCAGCGCAATATGATGATGAAGTGAAT	62
QY	21	ValLeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGluGluGluCysLysLysLeu	40
Ds	63	ATTTTGGCGTTGATTCTTCAACAGAAAGATGCAATGAGATACAAATGCAAAAAAGTTTA	122
QY	41	LysLysLysCysGluGluLeuThrGluAlaLysLeuAsnIleGluGlnValHisArgLys	60
Ds	123	GAAAAAATACCTGCGAAGAGTTGAAAAAGCATCACTAGCATGAAAAAGTACATCAATAAATG	182
QY	61	LeuLysGlyPheCysGluAspGly-----LysAlaAspThrLysCysLysGluLeu	77
Ds	183	CTTAAAGATTCTGTGAAATGGGAAAGCAAGTAAAGCAATATACAAATGTCAAGGTTTA	242
QY	78	LysAlaAsnIleGluLysCysThrThrIleLys--GlyLysLeuLysGluAlaIle	96
Ds	243	CAACCCAAAGTTACGGGGAATGTACAAATTTTAAACACAAAAGCTTAGCACAGCGTTA	302
QY	97	LysLysLysIleGlnIleIleThrAspLysAspCysLysGluAsnGluGlnCysLeu	116
Ds	303	ACAATCCA-----TCAGATGATATATGCAAAAGAGTGAACGACCAATGCTTA	350
QY	117	PheLeuGluGlyValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLys	136
Ds	351	TTTTTGGAGGAGCATGCG--CATATCTTGTTAAGAAAGTTGTAACAACTAAGAAATCTA	407
QY	137	CysTrpGlnLysLysAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSer	156
Ds	408	TGTTTACCGAAAAAAGTGAAGCGAGGTACAGAAAGAAAGTCTTTGAGGGCACTTCGTAGT	467
QY	157	AspLeuAsnGlySerValIleCysGluLysLysLeuLysGluIleCysProValIleGly	176
Ds	468	GATGTCATATAAACAAGAACACATGAATAAAAACTGAAGAGAGATTGGCCAGCTTTGCAG	527
QY	177	ArgGluSerAspGluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIle	196
Ds	528	AGGAAAGTAAATGAATTAACGAGCTTGTTGTTGAACAGAAAAAGAGCTGGAGAAATATT	587
QY	197	LeuIleGluLysAspLysLysCysGlyThrLeuLysThrAspValSerAlaIleGluGly	216
Ds	588	ATAAAGAAAAAGTAAAAAATGCACATCTTTAAAGCAAAATGTTGCACACGACCTTGA	647
QY	217	SerPheLysLysGluThrCysLeuGluLeuLeuGluGluGlnCysTrpPheTrpIleGlyAsn	236
Ds	648	AGTTTAAAAAAGAAATATGCTTGAATTACTTGAACAATGCTATTTTTCATGTGAAT	707
QY	237	CysGlyAspAspAspIleIleLysCysIleGluLeuGlyGlyLysCysGluGluGlnAsn	256

[illegible]

QY	616	ThrCysgluuewctvllvsaepvalvysaaprcysgluvalphelyslysaenlle	635
Db	1836	ACGtGTAGCGTGAAGTGAAGAAGTGAAGAAGCAAGGtGTAAAGTTCAAAATAATA	1895
QY	636	LysAlaSerTrilletiegupheuglvaasnaanthrasnlysllethrtthleuglu	655
Db	1896	AAACCTTCATATATATTAAGAAATTCCTGGAAAATAATCAATATAAATACACACTGGA	1955
QY	656	ArgAnCysPProSerTrpHisThrTrCysasnargPheSerProAnCysPProGlyLeu	675
Db	1956	AGAAATGTCCTCTTGCACTACTATATGCAATATGAAATATTCACCTAATTCCTCAAGGCTT	2015
QY	676	ThrIysgluaenserCysThrLysIleLysLysHisargGluPProPheTrLysArgLys	695
Db	2016	ACGAAAGAAATAGTTGTACAAAANATAGAGACATTGTAGCCGTTCTATAAAGAAAG	2075
QY	696	AlaLeugluabpalaleuLysValGluLeuGlnGlyLysLeuthraapLysSerLysCys	715
Db	2076	GCTTTGGAAGATGCTCTCAAAAGTGAAGCTTCAAGAAATAATTGACTGTAAATCTAAATGT	2135
QY	716	GluPProAlaLeuLysArgTrCysThrValAlaGlyAsnValAsenAlaSerIleSer	735
Db	2136	GAACCTGCATTGAAAAGATATGTACAGTGAAGCGGAAACGTAAATATATGCTCATAGT	2195
QY	736	GlyLeuCyLysAlaAenThrLysAspAsnSerGlyLysSeraspGluabpalargLys	755
Db	2196	GGCTTATGCAAAAGTTAACCCAAAGATATCTCTGGAAGAAGGTGATGAGATGCTAGAAAG	2255
QY	756	GluLeuCySgluLysLeuValLysGluValGluGluGlnCysLysValaleuPProThrGlu	775
Db	2256	GAACCTCTGTGAAGAAATTAGTGAAGAAAGATGGAAGAACAGTGCAAAGCATTTACCAACGAA	2315
QY	776	LeuGlyGlnPcoAlaAlaapLeuLysLysaspTrLysThrTrTrGluGluLeuLysLys	795
Db	2316	TTAGGACAAACCGGACGCTGATTTAAAAAAGATTATAGACATATGAGGAACTTAAGAA	2375
QY	796	ArgAlaGluGluAlaMetAsnLysSerSerLeuValLeuSerLeuIleLysLysAsnGlu	815
Db	2376	CGTCAGAGGAGCAATGAAACAACTCCAGCTTGTTGTTCACCTCATTAAGAAAAAGAA	2435
QY	816	SerPenValSerLysSerAsnSerLysAsnLysaspLysPenAlaValSerAsnGlyLeu	835
Db	2436	AGTATATGATCAAAAGATTAAGCAAAACAGGATTAAGATGCCGTTCAACCGGACTT	2495
QY	836	GlnAspThrThrLysHisValLysIleLeuArgArgGlyValLysAspValSerValThr	855
Db	2496	CAAGTACCAAAACATGTGAAAAATCTACGAGAGAGATTAAGATGATTCCTGTAA	2555
QY	856	GluLeuGluValAlaLysAlaPheAspLeuAlaAlaGluValPheGlyArgTrValAspLeu	875
Db	2556	GAATTTAGAGCTTAAAGCAATTTGATTTGGACAGCAAGATATTTGGAAATATGTAGATTTG	2615
QY	876	LysGluArgCysAsnLysLeuGlnSerAspCysArgLleLysGluAspCysLysAspLeu	895
Db	2616	AAAGAAAGATTAATTAATTTGCAATTCGATTCGCAAGATTTAAGAGGATTTGCAAAAGACTTA	2675
QY	896	GluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluValLys	915
Db	2676	GAAGAGATATCAAAAAAGATTAATAAGGCTTGTCGCAATCTGAAGCCTCTGGAGGTGAAG	2735
QY	916	ProHisGluThrValLtnGluSerThrThrThrThrThrThrThrThrThrValAla	935
Db	2736	CCGACCGAAACAGACAGAAAGTACACGCAACTACCAACAAACAAACAAACCCGTTGCC	2795
QY	936	AspProLysAlaThrGluCysLysSerLeuGlnThrThrAspThrTrpValLtnGlnThr	955
Db	2796	GATCGGAAGCAAGCAATGCAATCCTTACAGCAACACAGCACATCGGTTTACACACACA	2855
QY	956	SerThrHisThrSerThrSerThrIleThrSerThrIleThrSerLysIleThrLeuThr	975
Db	2856	TCGACACACACACACGCTTACTACTATACATCTACACATCACTCAACAAAATAACCTTACA	2915

QY 976 SerThrArgArgCysLysProThrLysCysThrThrGlyGluAspAspAlaGlyAsp 995
Db 2916 TCACGAGCGCATCAACCAACCAAGTGTACGACAGG-----GATGATGACGAAGAC 2969
QY 996 ValLysProSerGlyGlyLeuArgMetSerGlyTTPanValMetArgGlyValIleVal 1015
Db 2970 GTGAAGCCAACTGAAGGCTTGAAGGTCGCGGTGATGTGATGAGGGGGGTATGTA 3029
QY 1016 AlameValIleSerPheMetIle 1023
Db 3030 GCATGTGTTATTTCTGTCATGATT 3053

RESULT 5
AR438437 3084 bp DNA linear PAT 20-FEB-2004
LOCUS AR438437
DEFINITION Sequence 7 from patent US 6664053.
ACCESSION AR438437 GI:42663292
VERSION AR438437.1 GI:42663292
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3084)
AUTHORS Kovacs,J.A., Huang,S., Maaur,H., Fischer,S.H., Gill,V.J. and Mei,Q.
TITLE Identification of a region of the major surface glycoprotein (MSG)
gene of human Pneumocystis carinii
JOURNAL Patent: US 6664053-A 7 16-DEC-2003;
FEATURES
SOURCE Location/Qualifiers
1 . 3084
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2.27e-271 Length: 3084
Score: 4352.00 Matches: 826
Percent Similarity: 87.28% Conservative: 80
Best Local Similarity: 79.58% Mismatches: 106
Query Match: 80.43% Indels: 26
DB: Gaps: 6

US-10-654-416-14 (1-1023) x AR438437 (1-3084)

QY 1 MetAlaArgAlaValLysArgGlnAlaAlaGlyThrGlnAsnSerIleAspGluHis 20
Db 1 ATGCCGCGGGCGGTCCAGCGGACGCAAAAGCTGCAGATGATGATGATGAGAGCAT 60
QY 21 ValLeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGlnGluCysLysLysLeu 40
Db 61 GTTTAGCTTTGATTTAAAAAATAATGATTTAGAGATACAAAATGCAAAATCAAGTTG 120
QY 41 LysLysArgCysGlnGluLeuThrGlnAlaLysLeuAsnIleGluGlnValHisArgLys 60
Db 121 GAAACAATATTCGCAAAACATTTAAACAATGACGATTTAAATCCAGAAAAAGTTCCAGAAAA 180
QY 61 LeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAlaAsn 80
Db 181 TTAAAGATTTCTGTGATACGGAAACGAAATGAAATGAAATGTCAGATCTTAAAAACAA 240
QY 81 IleGluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLysLysIle 100
Db 241 GTCAATCAAAAATGATTAATTTCAAGGAAAATCTTCAACAGCGTCAGAAAAAAATTT 300
QY 101 GlnIleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeuPheLysGlu 120
Db 301 TCAGAAATTAACAGATGAGATTCGCAAAAAGATGAAACAATGCTATTTTGGAGGA 360
QY 121 ValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyrGlnLys 140
Db 361 GCATGTCCAAAGAACTTAAAGATGACTGCAATTAATTAAGAAATTAAGTGTATCAAAA 420
QY 141 LysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGly 160
Db 141 LysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGly 160

Db 421 GAAACGAAACAATGTGGCAGAAAGATCTTTTGAAGGCGCTTCGTGATCTCAATGAA 480
QY 161 SerValIleCysGluLysLysLeuLysGluIleCysProValMetGlyArgGluSerAsp 180
Db 481 ACAAGACATGTGAAAAAATCTGAAGAAAGTTGGCCGAAATTTAGAAAGAAAGCAT 540
QY 181 GluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIleLeuIleGluLys 200
Db 541 GAATTTACGAGGCTTTGCTTTTATCAAAAAACAACATGCCGTAAGTCTGTAAACAAAGA 600
QY 201 AspLysCysGlyThrLeuLysThrAspValSerIleAlaLeuGlySerPheLys--- 219
Db 601 AAAAGTAAATGTGATCTCTTGAAGAAAGATTGAAACACTTAAGAAATGAAATTG 660
QY 220 LysGluThrCysLeuGluLysLeuLysGluLysCysTyrPheTyrIleGlyAsnCysGlyAsp 239
Db 661 CGAGAAAAATGTCTACTATTTACTTGACCAATGTTACTTTTCACAGAGGAACTGCAAGA 720
QY 240 AspAsp-----IleIleLysCys 245
Db 721 GACAAATCAAAAGTCATTAATACTAATATATAAGACTGCAAAAGATATGTACAGAGTGT 780
QY 246 IleGluLeuGlyLysCysGlnGluGlnAsnIleAlaTyrMetProProGlyProAsp 265
Db 781 GATGAATTTACAGAAAGTGTGAAAAAATAATTTGTTATATGATCCAGAGATCCGAT 840
QY 266 PheAspProThrArgProGluAlaThrIleAlaGluAspIleGlyLeuGluGluPheTyr 285
Db 841 TTGCATCCAACTAAGCCAGAGCCTACCTAGCAGAGCAATAGGCGCTGGAAGACCTTAT 900
QY 286 LysLysValGluGluAspGlyValPheIleGlyLysAsnHisLeuArgAspAlaThrAla 305
Db 901 AAGAGGCGCAGAGAGATGAAATTTTGTGGAAGACAACTGTAAAGATGCAACAGCT 960
QY 306 LeuLeuAlaLeuLeuIleGlnAspSerSerLeuLysLysLysAspAspLysGluLysCys 325
Db 961 TTGTTGCACTACTT-----CTTAAGAAAAACCTTTAAAGAAAGAAATGT 1005
QY 326 GluGluAlaLeuGlnLysSerCysLysAsnProIleGluIleGluAlaLeuGluSerLeu 345
Db 1006 ATAAAGCCCTTAAAAAAAATCTGCAAAACCTCATGAACATGAGGCTTTAGAAAACTTA 1065
QY 346 CysLysLysAsnGlyLeuSerAsnAspGlyThrLysLysCysGluGluLeuGlnAsnAsp 365
Db 1066 TGTAAGAAAAATAACAAGATGATGGAACGAAAAAATGTGATGAACCTAAAGAAAGAT 1125
QY 366 IleAsnLysThrCysLysIlePheThrSerLysValThrAsnAsnArgLeuPheAspPro 385
Db 1126 GTTAACAACAATCTGTACAGCTTACATCAACAATTTCTTAACAAACCGTCTTACATTTCA 1185
QY 386 ThrLysGlyAsnAsnGluIleValGlyTTPGluGlyLeuProThrPheLeuSerAsnGlu 405
Db 1186 CCGATGAGA-----ATTGCGGAATGGGAAAAATTTACCGACATTTCTTAGAGAGA 1266
QY 406 AspCysAlaLysLeuGlnLysSerTyrCysPheTyrPheGluLysLysCysProAspGlyGlu 425
Db 1237 GATTTGCAAAAATCAAGATCTTATGCTTTATTAATAAGAAACTGTCCAGATGCATA 1286
QY 426 AsnAlaCysLysAsnIleArgAlaThrCysTyrLysArgGlyLeuAspAlaArgAlaAsn 445
Db 1297 GAAGCTTGTAATGATGAGAGGCGAGCTGTTAACAAGAGAGGCTGTGATGACGCGGCAAC 1356
QY 446 LysValLeuGlnGluAsnMetArgGlyMetLeuHisGlySerAsnLysSerTPLeuGlu 465
Db 1357 AGTGTTGTGCAAAAAATATAGCGGGGTATTCGTGTTCAAAATGAAAGTTGCTTAAG 1416
QY 466 LysPheGlnGlnGluLeuValLysValCysGluLysLeuLysGluAsnLysGlySer 485
Db 1417 GAGTTTCAACAAGATTAAGTAAAGATGTATGAGAGCTA---AAAGAAATTAAGAGAGT 1473
QY 486 PheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLysAlaAlaArgLeuLeu 505
Db 1474 TTCCCAACGATGAATAATTTGTTCTGTGTGTACAGCCAGCAAAAGCTGCACGATTAATT 1533

QY	506	ThrHisaApLeuAsgMetLeuThrIlePheLeuAtrGlnGlnLeuAapGlnLysArgAap	525
Db	1534	ACACACGATCATCAATAATGAGGGTTACCTTTTTCACACACAACTTGGATCAAAAGAGAT	15939
QY	526	PheProTArAepLysAasnCybLysGluLeuGlyAArgLysCybGlnAapLeuGlyLusAp	545
Db	1554	TTTCCGACAGATTAAGACTCGCAGAGAACTTAGGAAAAATGCCAAGATTTCAGAAAGAT	16553
QY	546	SerLysGluIleThrTTPProCysHisThrLeuGlnGlnCybAasnArgLysIleThr	565
Db	1654	TCAAAAGAAATTACATGCGCATGTCATACCTGAGCAGCAATGCATGCCTGGGGACT	17131
QY	566	ThrGluIleLeuLysGlnValLeuLeuAapGluVhiLysAapThrLeuLysAapGlnGlu	585
Db	1714	ACAGAAATTTTAACACAGGTTTTTATTTGGATGAACACAAAGATCTTTAAGAACCAAGAA	17739
QY	586	SerCybValLysTyrLeuLysGluLysCybAasnLysTTPSerArgArgGlyAAspAapArg	605
Db	1774	AGTTGTGTAAATACCTTAAGAAAGAGTAAATAAATGGTGTAGAAAGAGAGATGACCGT	18334
QY	606	PheSerPheValCysValPheGlnAasnAlaThrCysGluLeuMeValLysAapValLys	625
Db	1834	TTCTCTTTTGTATGTGTTTTCCAAAACCTACGTGTGAGCGATGTGTAAGAAGCCTGAA	18939
QY	626	AspArgCysGluValPheLysLysAasnIleLysAlaSerTyrIleIleGluPheLeuGlu	645
Db	1894	GACAGGTGTGAAGTATTCAAAATAATTAAGACCTCATATATTAATTCGATTTCTTGAA	19553
QY	646	AasnAnthrAasnLysIleThrThrLeuGluAtrGasnCybProSerTTPHisThrTyrCys	665
Db	1954	AATAATACAAATAATAATAACACACTGGAAATAATTCCTCTTGGCATAACGTAATGC	20134
QY	666	AasnArgPheSerProAasnCybProGlyLeuThrLysGluAasnSerCybThrLysAlaLys	685
Db	2014	AATAGATTTTCACCTAATTTGTCAGGCTTACGAAGAAGATAGTTGTACAAAATCAAG	20739
QY	686	LysHisArgGluProPheTyrLysArgLysAlaLeuGlnAapAlaLeuLysValGluLeu	705
Db	2074	AAGCATTTGAGACCCGTTCAATAAAGAAAGGCTTGGAAGATGCTCTCAAGTAAAGCTT	21339
QY	706	GlnGlyLysLeuThrAapLysSerTyrCysGluProAlaLeuLysAArgTyrCysThrVal	725
Db	2134	CAAGGAAATATGACGTATTAATCTTAAGTGAACCTGCATGGAACAGATATTGTACATA	21939
QY	726	AlaGlyAasnValAasnAasnAlaSerIleSerGlyLeuCybLysValAasnThrLysAapA	745
Db	2194	GGGGAAACGTAAATTAATGCGCATACGTGCTTATGCAAAGCACTAACCCAGAGATAC	22553
QY	746	SerGlyLysSerAapGluAapAlaArgLysGluLeuCybGluLysLeuValLysGluVal	765
Db	2254	TCTGAAAGAGATGAGAGATGCTTGAAAGAACTCTGGAATACTAGTGAAACAGATGTG	23134
QY	766	GlnGlnGlnCybLysAlaLeuProThrGlnLeuGlnGlnProAlaAlaAapLeuLysLys	785
Db	2314	GAAGAACAGTGCAGAACGATTAACCAACAGATTAGACAAACCGGCGCTGATTAATAAAA	23739
QY	786	AepTyrLysThrTyrGluGluLeuLysLysArgAlaGlnGlnAlaMeLAsnLysSerSer	805
Db	2374	GATTTATTAAGACATATAGAGAACTTAAGAAACGTGACGAGAGAAAGCAATGAACAGTCCAGT	24339
QY	806	LeuValLeuSerLeuIleLysLysAasnGlnSerAasnValSerLysSerAasnSerLysAasn	825
Db	2434	CTTGTTTGTCTACTATTAAGAAACGAAAGTAAATGTATCAAAAAGTAATGCAAAAAC	24939
QY	826	LysAapLysAasnAlaValSerAasnGlyLeuGlnAapThrThrLysHisValLysIleLeu	845
Db	2494	AAGATTAAGAAAGCCGTTCAAAAGGAGCTTCAAGTACCAACAAACATGTAATAATCTA	25539
QY	846	ArgArgGlyValLysAapValSerValThrGluLeuGluAlaLysValAapPheAapLeuAla	865
Db	2554	CGGAGAGGATTAAGATGTAATCCGTACAGAAATTAGAAAGCTTAAGCAATTTGATTTGGCA	26134

QY	866	AlaGluValPheGlyArgTrValAspLeuLeuGluArgCysAsnLeuLeuGluSerAsp	885
Db	2614	GCAGAAAGATTTGGAAAGATGTAATTTGAGGAAAGATGTAATTAATTTGAAATCGAT	2673
QY	886	CysArgIleuLeuGluAspCysValAspLeuGluGluValCysValCysIleAsnLeuAla	905
Db	2674	TGCAGATTTAAGGAGGATTCGAAAGACTTAGAGAGAGATATCGAAAAAGATTAAATAGGCT	2733
QY	906	CysArgAsnLeuLeuProLeuGluValIleProHisGluThrValIleArgLeuSerThrThr	925
Db	2724	TGTGCAGATCTGAAGCCCTCTGAGAGTGAGGCGGACCGAAACAGTACAGAAAGTACAACG	2793
QY	926	ThrThrThrThrThrThrThrValAlaAspProLysAlaThrGluCysLeuSerLeu	945
Db	2794	ACAATCTACACACAAACAAACAAACCGTGGCCGATCCGAAGGCACGGAATCCAATCCTTA	2853
QY	946	GlnThrThrAspThrTrpValIleGlnThrSerThrHisThrSerThrSerThrIleThr	965
Db	2854	CAGACACAGACACATGGGTTCACAGACATCGACACACACACAGCAGCTTACTATACACA	2913
QY	966	SerThrIleThrSerLysIleThrLeuThrSerThrArgArgCysLysProThrLysCys	985
Db	2914	TCTACATCATCATCAAAAATTAACATTGATGCATCAACGAGCGATCAACCAACCAAGTGT	2973
QY	986	ThrThrGlyGluGluAspAspAlaGlyAspValLysProSerGluGluLeuArgMetSer	1005
Db	2974	ACGACAGGG-----GATGATGACGAAAGACGTGAAGCCAAAGTGAAGGCTTGAGGGTACG	3027
QY	1006	GlyTrpAsnValMetCysGlyValIleValAlaMetValIleSerPheMetIle	1023
Db	3028	GCGTGGAATGTGATGAGGGGGGTGATGTAGCAATGATTATTTCGTTCAATGATT	3081
RESULT 6			
LOCUS	AF033209	3083 bp DNA linear	PLN 10-SEP-1998
DEFINITION	Pneumocystis carinii f. sp. hominis clone HUMSG14 major surface glycoprotein (MSG) gene, partial cds.		
ACCESSION	AF033209		
VERSION	AF033209.1	GI:3560514	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			</

QY	507	H1SAAPLEUAIGMELVETHRI1LEHELEUARGINGINLEUAPRGINLYAARGAPHE	526
Db	1536	CACGATCATCAAAATGAGAGTTACCTTTTATGACAAACAATTGGATCAAAAGAGATTTT	1595
QY	527	PROTHRAAPLYASANCYSELG1ULEUOLYARGLYSCYSG1NAEPLEUG1YGLUABOSER	546
Db	1596	CCGACAGATTAAAGACTGCAGGAACCTTAGGAAAATAATCCCAAGATTTAGAAAGATTCA	1655
QY	547	LYSG1ULLETHRTTPROCYSH1ETHREUGLUGINGINCYBAANXLEUGLYTHRTHR	566
Db	1656	AAAGAAATTACATGGCCATGTCAATCATGGGACGACAAATCAATCGCTTGGGGACTCA	1715
QY	567	GIULIELEUYSGINVALLEULEUABPG1UNH1SLYSAAPRTIRLEUYAAPG1G1LSER	586
Db	1716	GAATATTTAAAGCAGCTTTTATTGGATGACCAAAAGATACTTTGAAGACCAAGAAAT	1775
QY	587	CYSVALYSTRYLEUYSGLULYSCYBAANLYSTRPSEARGARG1AASPAPRPH	606
Db	1776	TGTGTAAATACCTTAAGAAAGAGTAAATATGTCTTAAGAAGAGATGACCGTTTC	1835
QY	607	SERPHEVALCYSEVALPHEGLNABNALATHRCYSG1LEUETVALYSEAPVALYSEAP	626
Db	1836	TCTTTGTATGTTGTTCCTCAAAACCCCTACGTGTGACGTAGGTAAAGACGTGAAGAC	1895
QY	627	ARGCYSG1ULVALPHELYSELYASAN1LELYSLASERTYRIELLEG1LPHLEUG1LBN	646
Db	1896	AGGTGTGAAGTATTAABAAAAATTAABAGCTTATATATTAATGAATTTCTGAAAT	1955
QY	647	ASANTHRAENLYSL1ETHRTHREUG1UARXANCYSPROSETRPH1ETHRYCYBAN	666
Db	1956	AATACAAATTAATAATNACAACCTGGAAAAGAAATGTCCCTCTGGCATACGATTCGAT	2015
QY	667	ARGPHESEPROABENCYSPROG1YLEUTHRYLYSG1UBANSECYSTHLYSL1ELYLYS	686
Db	2016	AGATTTTACACCTTAATGTCCAGGCCCTTGCGAAAGAAATAGTTGTACAAAATCAAGAG	2075
QY	687	H1SARG1UNPROPHETRYLYSEARGLYVAL1LEUG1UABAPAL1ELEUYVAL1G1LEUG1N	706
Db	2076	CATTGTGACCGCTTTTAATAAAGAAAGGCTTGGAAGATGCTCTAAGATAGAGCTTCA	2135
QY	707	G1LYLYLEUTHRAAPLYSESELYSCYSG1UPROAL1ELEUY1SARGTRYCYSTHRYVAL1A	726
Db	2136	GGAAATTTGACTGATTAATCTTAATGTGAACCTGCATTTGAACAGATTATGTACAGTACG	2195
QY	727	GLYASENVALAENNALASERT1LESERGLYLEUCYSLY1SALABENTHLYSEAPBANSE	746
Db	2196	GGAAACGTAATTAAGCGCTCATCATGTGGCTTATCCAAAGCTAACCAACCAAGATTAACT	2255
QY	747	G1LYLYSESEARAPG1AASPALAARGLYSG1ULEUCYSG1ULYLYLEVAL1LYSG1UVAL1U	766
Db	2256	GGAAAGATGATGAGAGATGCTTAGAAGGAACTCTGTGGAATCTAGTAAAGAAAGTGA	2315
QY	767	GIUG1INCYLYVAL1LEUPROTHRG1LUGING1GINPROAL1A1AASPLEUYSLYSEAP	786
Db	2316	GAACAGTCCAAAGCATTACCAACAGATTAGGACAAACCGGACGTGATCTTAATAAAGAT	2375
QY	787	TYRLYSTHTRYTGLUG1ULEU1LYSARG1AAG1UG1U1AMEASMLYSESER1EUN	806
Db	2376	TATAAGACATATGAGGAACCTTAAGAAAGCTGCAGGAAGCAATGACAAAGTCCAGTCTT	2435
QY	807	VALLEUSERLEU1LELYLYSEANGLUSER1AENVALSERTLYSESEARPSERTLYSEANLYS	826
Db	2436	GTTTGTCACTCATTTAGAGAAAACGAAAGTATGTATCAAAAAGTAAATAGCAAAAACAG	2495
QY	827	ASPLYASAN1AVAL1SERASNG1YLEUG1NBPRTHTHRYSH1SEVALYSL1ELEUARG	846
Db	2496	GATTAGAATGCCGTTTCAACGGACTTCAAGATTAACAACAAACATGTGAATAATCTACG	2555
QY	847	ARGGLYVALYSEAPVAL1SERVAL1THRG1LEUG1U1ALYVAL1PHEAPLEUAL1A	866
Db	2556	AGAGGAGTTAAAGAGTATCCGTAAACAGATTAGAAAGCTTAAAGCACTTTGATTTGGACCA	2615

QY	867	GIUVALPHEGJYAGYLYRVAIAspLeuLYGLuRGYsAsnLYsLeuGLuSERAspCYs	886
Db	2616	GAAGTATTGTGAAATATGTAATTTGAAGAAAGATGTAAATAATTGCAATCAGATTGC	2675
QY	887	ArgIleLYsGIuAspCYsLYsAspLeuGLuGLuVALCYsLYsLYsIleAsnLYsAlaCYs	906
Db	2676	AGAATTAAAGGAGGATTCGCAAGACCTTGGAAGAGTAATGCAAAAAGATTAAATTAAGGCTTGT	2735
QY	907	ArgAsnLeuLYsPLeuGLuGLuVALCYsLYsProHISGluThrValThrguSerThrThrThr	926
Db	2736	CGCAATCTGAAGCCTCTGGAGGTAAAGCCGACAGAAACAGTGACAGAAAGGTACAAACACA	2795
QY	927	ThrThrThrThrThrThrThrThrValAlaAspProLYsAlaThrgIuCYsLYsSerLeuGLn	946
Db	2796	ACTACAACAACAAACACACCGCTTGCGCATCCGAAGCAACGAAATGCAATCTTAAAG	2855
QY	947	ThrThrAspThrTPpValThrgInThrSerThriThriSerThrSerThrIleThrSer	966
Db	2856	ACAACAGACACATGGGTACACAACACATCGACACACAAAGCAGTCTACTATTCACATCT	2915
QY	967	ThrlIeThrSerLYsIleThrLeuThrSerThrTrgArgCYsLYsProThrLYsCYsThr	986
Db	2916	ACCATCATATPAAAAATPACATTACATTCACAGAGCGCATGCCAAACCAACCAATGTACG	2975
QY	987	ThrgIuGLuGLuAspAspAlaGLyAspValLYsProSerGLuLYsLeuArgMetSerGLy	1006
Db	2976	ACAGGG-----GATGATGTCAGAAAGCGTGACGCAAGTCAAGCCTTGAGGCTGAGCGGG	3029
QY	1007	TrpAsnValMetArgLYsValIleValAlaMetValIleSerPheMetIle	1023
Db	3030	TGGAAATGTAGTGAAGGGGGGTGATGTAGTGCATGTGTATTTCGTTCAATGATT	3080
RESULT 7			
LOCUS	AR438436	3090 bp	DNA linear PAT 20-FEB-2004
DEFINITION	Sequence 5 from patent US 6664053.		
ACCESSION	AR438436		
VERSION	AR438436.1	GI:42663291	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 3090)		
AUTHORS	Kovacs,J.A., Huang,S., Maaur,H., Fischer,S.H., Gill,V.J. and Mei,Q.		
TITLE	Identification of a region of the major surface glycoprotein (MSG)		
JOURNAL	gene of human Pneumocystis carinii		
FEATURES	Patent: US 6664053-A 5 16-DEC-2003;		
source	Location/Qualifiers		
	1..3090		
	/organism="Unknown"		
	/mol_type="genomic DNA"		
ORIGIN			
Alignment Scores:			
	Pred. No.:	1..66e-237	Length: 3090
	Score:	3827.00	Matches: 728
	Percent Similarity:	81.54%	Conservative: 120
	Best Local Similarity:	70.00%	Mismatches: 164
	Query Match:	70..73%	Indels: 28
	DB:	6	Gaps: 8
US-10-654-416-14 (1-1023) x AR438436 (1-3090)			
QY	1	1 MetAlaArgAlaValLYsArgGLnAlaAlaGLYThrgInAsnSerIleAspGLuGLuHis	20
Db	1	ATGGCGCGGGCGGTCAAGCGCGGCAAAAGGTCCACAGAAATGACATTGATGAGAGCAT	60
QY	21	ValLeuAlaLeuIleLeuLYsGLuAspGLYLeuSerGLuGLuGLuGLuCYsLYsLYsLeu	40
Db	61	GTTTTAGCTTGATTTTAAATAAAATGATTTAAGAAGTACAAAATGCAAACTTAAGTTG	120
QY	41	LYsLYsTYrCYsGLuGLuLeuThrThrgValaLYsLeuAsnIleGLuInValIleArgLYs	60

Qy		76	GIuVaLGIuGIuGIuGNCyVbYbAlaLeuProThrGIuLeuGIuGIuProAlaLaLaAspLeu	783
Db		2314	GAGGIGGAGCAGCAATGCAGATATTACCAACAGATTAAACAGAGCTGGAAAAAGTCTA	2373
Qy		784	LYsLYsAspTYrLYsThTYrGLuGIuLeuLYsLYsATGAlaGIuGLuAlaMeCAsnLYs	803
Db		2374	AAAAAAGATGTTAAGACATATAGAGGAATTAAAGAAAGCGCAAAAAAGCATGACAG	2433
Qy		804	SeSerSerLeuValLeuSerLeuIleLYsLYsAngLusSerAsnValSerLYsSerThsSer	823
Db		2434	TCCAGCCTTGTTTATCACTGTTTAAAGAAAAACGAAGTAATATACATCGAAAAATATATAGC	2493
Qy		824	LYsAsnLYsAspLYsAsnAlaValSerAsnGLuLeuGIuAspThrThLYsHisValLYs	843
Db		2494	AAAAACAAGATTAAGAAATGCTGTTCAACGCACTCAAGATATCCACAAAATATATGTGAA	2553
Qy		844	IleLeuATGATGAGLYsValLYsAspValSerValThrGIuLeuGIuAlaLYsAlaPheAsp	863
Db		2554	ATTACTCAAAAGAGAGATTAAAGAGGCACTTGTTAAACGAATCTGAAGCCAAAGCATTTGAT	2613
Qy		864	LeuAlaAlaGIuValPheGIuYArgTYrValAspLeuLYsGIuArgCYsAsnLYsLeuGIu	883
Db		2614	TTGGCAGACAGAAAGCTTTGAAAGATATATACCTTGAAGAAAAAATGGAAGAAATTGACT	2673
Qy		884	SeSerSpCYsAAGIleLYsGIuAspCYsLYsAspLeuGIuValCYsLYsLYsIleAsn	903
Db		2674	TCCGATTTGCGGATTTAAAGACGATTTCCAGTGGTTTAAAAAGAGTGtGGAAGAGATTGAG	2733
Qy		904	LYsAlaCYsATGAsnLeuLYsProLeuGIuValLYsProHisGLuThValThrGLuSer	923
Db		2734	AAGACATGTCACGATCTGAACCTCTGAGGtGAAAGTCGCATGAATGTCTACAGAAAGC	2793
Qy		924	ThrThrThrThrThrThrThrThrThrThrValAlaAspProLYsAlaThrGluCYsLYs	943
Db		2794	ACAACGACGACCAACACAGCAAAACGACCGTTACCGATCCGAAGCCAAACAGATGCAA	2853
Qy		944	SeLeuLeuGIuThrThrAspThrThrValThrGIuThrSerThrHisThrSerThrSerThr	963
Db		2854	TCTTTACGACAAACGATACATGGGTTACACAGATGACACACACAAAGACCGTTAC	2913
Qy		964	IleThrSerThrIleThrSerLYsIleThrLeuThrSerThrAspArgCYsLYsProThr	983
Db		2914	ATCACATTTACATCATCATTCAAAAATTAACATTGACATCAAGAGCGATGCMAACCAAC	2973
Qy		984	LYsCYsThrThrThrGLuGIuAspAspAlaGLyAspValLYsProSeSerGLuGLyLeuArg	1003
Db		2974	AAGGTACGACAGAGG-----GATGAACACAGAGACGTAACCGAGTGAAGATTGAAG	3027
Qy		1004	MeSeSerGLyThrPheAsnValMeCArgGLyValIleValAlaMeCValIleSerPheMeTle	1023
Db		3028	ATGAGTGGGTGAGAGCGGTATGAGGGGGGTGATAGTAAAGATGTTATTCGTTCATGATT	3087
RESULT 8				
AF033208				
LOCUS		3089 bp	DNA	linear
DEFINITION				PLN 10-SEP-1998
				Pneumocystis carinii f. sp. hominis clone HUMS011 major surface
				glycoprotein (MSG) gene, partial cds.
ACCESSION				
AF033208				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
				Pneumocystis jirovecii
				Pneumocystis jirovecii
				Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
				Pneumocystidaceae; Pneumocystis.
REFERENCE				
1				(bases 1 to 3089)
AUTHORS				Garde,T.R. and Stringer,J.R.
TITLE				Molecular characterization of clustered variants of genes encoding
				major surface antigens of human Pneumocystis carinii
JOURNAL				Infect. Immun. 62 (8), 3092-3101 (1994)
MEDLINE				94314421
PUBMED				751806
REFERENCE				2 (bases 1 to 3089)
AUTHORS				Mei,Q., Turner,R.E., Social'V., Klivington,D., Angus,C.W. and

TITLE	Kovacs, J. A.
JOURNAL	Characterization of major surface glycoprotein genes of human
MEDLINE	Pneumocystis carinii and high-level expression of a conserved
PUBMED	region
98380374	
9712777	
REFERENCE	3 (bases 1 to 3089)
AUTHORS	Mei, Q., Turner, R., Social, V., Livingston, D., Angus, C.W. and
TITLE	Kovacs, J. A.
JOURNAL	Direct Submission
3089	Submitted (07-NOV-1997) Critical Care Medicine Dept., National
1662	Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD
1662	20892-1662, USA
FEATURES	Location/Qualifiers
SOURCE	1..3089
gene	/organism="Pneumocystis jirovecii"
CDS	/mol_type="genomic DNA"
	/specific_host="Homo sapiens"
	/sub_species="hominis"
	/db_xref="taxon:42068"
	/clone="HUMSG11"
	/note="derived from HIV-infected human with P. carinii
	pneumonia"
	<1..3089
	/gene="MSG"
	<1..3089
	/gene="MSG"
	/note="surface antigen"
	/codon_start=3
	/product="major surface glycoprotein"
	/protein_id="AAC34971.1"
	/db_xref="GI:3560513"
	/translation="ARAVKRAKAGNSIDSEHVALLIKNGLEDTCCKTXLEEXCK
	TLTNAGINPEKVEHLEKDFCDNGKRNKCCODLRKNVOKCIFKQGLQTAQKIKSEL
	TDEDCCKNQEQGLKFGACPTLKDCCNCKLRNNCYOKERNVAVELRLARDLNET
	KTECKELKYCPKLEBSDELTELCLYKTKTCVSLVTKGSKCDTLEKVEALDKNE
	LRCKCLILEQCYFHRNGCEGDKSKCKRNPKCKEYVPCDEBLAEKCKENIYVNH
	GSDFDTPKEPPTLAIEDIGLEELYKRAEBDI FGRGHVRAVATLALLKTLKKEC
	IKALKKNCENPHEHEALENLCKENKSSDPTKCDLEKDVNTCTLSITLILNRLY
	ISPDGLAEWKEHLPTPLSDEDCALAESYCFYKETCPDVKACNVBAACYKRGIDARA
	NSVLKONRBLHSGNKDMLKKRQELALCKELKGNKGSFSDLELPVLCTOPAKAR
	LILTHHOMRYIFLRQOLDKRPDPDKCELEKGRKODLKGSKELTMRPHTEQCN
	RLGLETILKQILIDHEKDTLKHENCAKTYLKRCKHRSRRGDRFSFVCFQVATCEL
	MVDVODRCKIFEEFNQASDINDSLKNOIKASANI CPSMRYCDRLPLNCPDLK
	GKFFCQDLKKCYCEPFYKRYVLEALVELKLENSITKECPALERYCTVLADVNMSI
	SSLCNDTESKTKKADKNKVRKTLCLKLVVEVQOCVLPTLELEKSLKDKVKTVE
	ELKEPRKAKNKSILVSLVKNKSNMSTKSNKSKNDKNVSNGLDPTTKYKILRGV
	KHALVTRSEAKADPLAAEVGRVYDLKEKEKTLSPDGLIDDDGDLKNGCKI EKTH
	DLKPLEKSHIEIVTESITTTTTTTTTTVDTPKATCEKSLQTTDVTWOTSTHSTSTT
	STTSKTLITSTSRCKPTKCTGTDEAGDVKPSBGLKMSGWSVMRGVIAVMVIFSMI"
ORIGIN	
Alignment Scores:	
Pred. NO.:	3.48e-237
Score:	3822.00
Percent Similarity:	81.52%
Best Local Similarity:	69.97%
Query Match:	70.63%
DB:	8
	Gaps: 8
US-10-654-416-14 (1-1023) x AFG033208 (1-3089)	
QY	2 AAlarqAlaVallysaAaGAlaAaAGlYThrClnaamSerIleaERgIugLuhIeVal 21
Db	3 GCGGGGGCGGTCAAGCGGGCGGCAAAAGGTGCACAGAAATGCAATTATGAGAGCATGTT 62
QY	22 LeuAlaLeuIleuLysGluAspGlyLeuSerGluGlnGluLysGlyLysGlyLys 41
Db	63 TTACCTTGAATTTTAAAAAATGATTTTGAAGATGCAAAATGCAAAACCTAAGTTGAA 122
QY	42 LysTTCyGselngLueuthrGluAlaLysLeuLamIleGluGlnValHisAqVlyLeu 61

Db 123 GAATATTGCAAAACATTAAACAAATGACAGATTAAATCCAGAAAAAGTTCCAGAAAAATTA 182
 Qy LygIlyPheCysGluAaspGlyValAlaAspThrIlySersGluLeuValAlaAsnIle 81
 Db 183 AAAGATTTCGTGTATACCGGAAACGAAATGAAATGTCATTAATAAACAAAGTC 242
 Qy 82 GluIlySersCysThrThrIleIlyGlyIlyLeuIlyGluAlaIleIlySlyIleGln 101
 Db 243 AATCAAAAATGCAATTAATTCAAGGAAAACTTCAACAGCTGCGGAAAAAATTTCA 302
 Qy 102 IleIleThrAspIlyAspCysIlyGluAengIuGlnGlnCysLeuPheLeuGluIlyVal 121
 Db 303 GAATTAAACAGATGAGATTGCAAAAAGATGAAACAACATCCATATTTTTCAGAGGAGCA 362
 Qy 122 CysSerIlySgluLeuIlyAspAspCysAsnThrIleuArgAsnIlySersGlyIlyLeu 141
 Db 363 TGTCCAACAGAACTTAAGATGACGCAATTAATTAAGAAATTAATCTGTTATCAAAAAAGAA 422
 Qy 142 ArgAspIlyValAlaGluGluValIleuLeuArgAlaLeuArgSerAspLeuAengIlySer 161
 Db 423 CGGAACAATGTCGAGAAAGATTCTTTGAGGCGCTTCGTGTGATCTCAATGAAACA 482
 Qy 162 ValIleCysGluIlyIlyLeuIlyGluIleCysAspValMetGlyArgGluSerAspGlu 181
 Db 483 AAGACATGTGAAAAAAGCTGAAAGAGTTGCCGAAATTAAGAAAGGAAAGCATGAA 542
 Qy 182 LeuThrAsnLeuIlySleuAengIlyGluIlyThrCysIlyAsnIleLeuIleGluIlyAsp 201
 Db 543 TTAACGAGACTTGTCTTATCAAAAAACAATGCTTAAGCTTGTGAACAAAAAGAAAA 602
 Qy 202 IlyIlySersCysGlyThrLeuIlyThrAspValSerAlaIleuGlySerPheIlySers 220
 Db 603 AGTAATATGTATATCTTTGAAAAAAGAGTTGAAGAGACTTAAGAAAGAAATTCGCA 662
 Qy 221 GluIlyThrCysIleuGluLeuLeuGlnGlnCysAspThrIleGlyAsnCysGlyAspAsp 240
 Db 663 GAAAAATGTCTACTATTACTTGACCAATGTTACTTTCACAGAGGAACTGTGAAGGAGAC 722
 Qy 241 Asp-----IleIleIlyCysIle 246
 Db 723 AATCAAAAGTCAATTAACCTTAATTAATTAAGCTGCAAAAGAAATGTAACAGAGTGCAT 782
 Qy 247 GluIleuGlyIlyIlySersGlnGluIlyAsnIleAlaIlyIleCysProGlyIlyAspPhe 266
 Db 783 GAATTACAGAAAAGCTGTGAAAAAATTTGTTATATGATCCAGATCCGATTTCC 842
 Qy 267 AspProThrArgProGluAlaThrIleAlaGluAspIleGlyIlyGluIlyPheIlySers 286
 Db 843 GATCCAATTAAGCCAGAGCTTACCTAGCAAGAGACATAGGCGTGGAAAGACTTTATAG 902
 Qy 287 IlyValAlaGluIlyAspGlyValPheIleGlyIlyAsnIleLeuArgAspAlaThrAlaLeu 306
 Db 903 AGGCAAGAAAGATGAAATTTGTTGGAAGACAACTGTMAAGATGCAACAGCTTTG 962
 Qy 307 LeuAlaLeuLeuIleGlnAspSerSerIlyIlySlyIlyAspAspIlyGluIlyCysGlu 326
 Db 963 TTGGCACTACTT-----CTTAAGAAAACCTTAAGAAAAGAAAGATGATATA 1007
 Qy 327 GluAlaLeuGlnIlySersCysIlyAsnProIleGluIleGluAlaLeuGluSerIlyCys 346
 Db 1008 AAAGCCCTTAATAAAAAACCTCATGAACATGAGGCTTGAATAAATCTATGCT 1067
 Qy 347 IlyIlyAsnGlyIlySerSerAsnAspGlyThrIlySersCysGluIlyLeuGlnAsnAspIle 366
 Db 1068 AAGGAAAAATTAACCAAGTATGATGAACGAAAAATGTGATGAACGAAAAAGATGTT 1127
 Qy 367 AsnIlyThrCysIlyIlePheThrSerIlyValThrAsnAsnArgLeuPheAspProThr 386
 Db 1128 AACCAAACTTATCAAGCTTACATCAACATTTCTTAATAAACCGCTTTACATTTCCACT 1187
 Qy 387 IlyGlyIlyAsnAsnGluIleValGlyTTrpGluIlyLeuProThrPheLeuSerAsnGluAsp 406
 Db 1188 GATGGA-----ATTGCGGAATGGGAAAAATTAACCGACATTTCTTAAGATGAAGAT 1238

Qy 407 CysAlaIlySleuGluSerIlyCysPheIlyPheGluIlySersCysProAspGlyGluAsn 426
 Db 1239 TGTCCAAAATCAAGAACTTATGCTTTTATTAATAAGAACTTGTCCAGATGTCAAAGAA 1298
 Qy 427 AlaCysIlyAsnIleArgAlaThrCysIlyIlyAspGlyIlyAspAlaArgAlaAsnIlyS 446
 Db 1299 GCTTGATGAATGAGGAGCGCTGTTATTAAGAGAGCGCTTGATGACGCGGCAAAACAGT 1358
 Qy 447 ValIleuGlnIlyAsnMetArgIlyMetLeuIleGlySerAsnIlySerTrpLeuGluIlyS 466
 Db 1359 GTGTTGCAAAAAATATGCGAGGTTATGTCATGCTCAAAATTAAGATGCGCTTAAGAAA 1418
 Qy 467 PheGlnGlnIlyLeuValIlyValCysGluIlySersIlySgluAsnIlyGlySerPhe 486
 Db 1419 TTTCAACAAGATTAAGCAAAAGTATGAGAAATG---AAAGAAATTAAGAAAGTTTC 1475
 Qy 487 SerAsnAspGluLeuPheIleLeuCysValGlnProAlaIlyValAlaArgLeuIlyThr 506
 Db 1476 TCGAAACGATGAATGTTGTTCTGTGTATCAACCAAGCAAAAGCAGCAGCATTTACTTACA 1535
 Qy 507 HisAspLeuArgMetIlyThrIlePheLeuArgGlnIlyLeuAspGluIlyAspAspPhe 526
 Db 1536 CATCACCATCAAAATGAGATTAATCTTTTACGACCAACACTGATCAAAAGAGAGATTTT 1595
 Qy 527 ProThrAspIlyAsnCysIlySgluLeuGlyValArgIlySersGlnAspLeuGlyIlyAspSer 546
 Db 1596 CCGACATTAAGACTCGAAGAAATTAAGGAGAAATCCCAAGTTTAAGAAAGATTTCA 1655
 Qy 547 IlyGluIleThrTrpProCysHisThrLeuGlnGlnCysAsnAsnArgLeuGlyIlyThr 566
 Db 1656 AAAGAAATTCATGCGCATGATACATGACAAACAGCAATGCAATCCCTTGAGGATTAACA 1715
 Qy 567 GluIleLeuIlySgluValIleuLeuAspGluHisIlyAspThrIlyIlyAspGlnIlySer 586
 Db 1716 GAAATTTTAAACGATTTATGATGAACACAAATGCTTTGAAAGATCATGAAGAAC 1775
 Qy 587 CysValIlySerIlyLeuIlySgluIlyCysAsnIlySerIlyArgIlyAspAspAspPhe 606
 Db 1776 TGTCCAAAATTTTAAAGAAATGCCATTAATGCTTCAGAGGCGTGAATGATCTTTT 1835
 Qy 607 SerPheValCysValPheGlnAsnAlaThrCysGluLeuMetValIlyAspValIlyAsp 626
 Db 1836 TCTTTGATGATGTTTCCAAAACGCTACATGTGAGCTGATGTAAGAAAGCTGCAAGAT 1895
 Qy 627 ArgCysGluValPheIlyIlyAsnIleIlyAlaSerIlyIleIleGluPheLeuGluAsn 646
 Db 1896 AGGCAAAATATTCGAAAGAAATATGCAAGCATCATATTAATGATTCCTTAATAAAA 1955
 Qy 647 AsnThrAsnIlySerIleThrIleuGluArgAsnCysProSerTrpHisThrIlyCysAsn 666
 Db 1956 AATCAAAATTAAGCAAGATGACGAGCAAAATTTGTCCTCATGAGCATCCATCTGCAT 2015
 Qy 667 ArgPheSerProAsnCysProGlyIlyLeuThrIlySers---GluAsnSerCysThrIlyS 685
 Db 2016 AGATTTTACCAATGATGCTCGATCTTAAGAAAGAAAACTTCTGTCAAAAATCTTAAA 2075
 Qy 686 IlyHisArgGluProPheIlyIlyValArgIlyAlaLeuGluAspAlaLeuIlyValGluLeu 705
 Db 2076 AAATATTGCGAACCATTTCAAAAAGAAAGTTTAAAGATGCTTTAAAGTGAAGCTT 2135
 Qy 706 GlnIlyIlySersIlyAspIlySersCysGluProAlaLeuIlyAspIlyCysThrVal 725
 Db 2136 CGAGGAAATTTAAGTAATTAATTAATTAATGAACTGATTAAGAAATATTTGATACATA 2195
 Qy 726 AlaGlyIlyAsnValAsnAsnAlaSerIleSerCyllyLeuCysIlySlyAlaAsnThrIlyAspAsn 745
 Db 2196 TTGAAGACGTAATTAATGCGTCAATCAGCAGTTATGTAAGATTAATACGAAAGTAAA 2255
 Qy 746 SerCyllyIlySerAsp---GluAspAlaArgIlySgluLeuCysGluIlySleuValIlyGlu 764
 Db 2256 ACTTAAGAAAGCGGTAATTAATAAATGTTAGAAAGAAAGCTTTGTCTTAATAATTAAGTGAAGAG 2315

[illegible][illegible]

Db 512 GCATGCG---CCTGATCTTGTAGAGATTGTAGCAAACTAAGCAATCTATGTTACAGAAA 568
 Qy 141 LysArgbAspValaIaGluValIleuLeuArgaIaLeuArgSerAspLeuAsnGly 160
 Db 569 AAACGTGAAGAGTAGACAGAAAGAGCTTTTGAAGGCACTTCGTGGGATTTTGAAGAA 628
 Qy 161 SerValIleCysGluValyLeuLeuArgGluIleCysProValMetGlyArgGluSerAsp 180
 Db 629 AAAACGTAAATCGAAAAGAAAATAAGAGTGTGGCCCAAAAATATAGGCCAAGAAAGTAT 688
 Qy 181 GluLeuThrAsnLeuCysLeuAsnGlnIlyeGluThrCysValyAsnIleLeuIleGluIys 200
 Db 689 GAGTTGACATTTGTGTGTCTGTGATCAGAAAGAAACATGCAAGATCTTATCTACCGCGA 748
 Qy 201 AspIlyAspCysGlyThrLeuLeuThrAspValSerIaIaLeuGluSerPheLeuIys 220
 Db 749 GACAAAAGTGTATACCTTGAGAGAGATGTTAAAAAGCATTTGAAGAACAAAAACAAC 808
 Qy 221 -----GluThrCysLeuGluIleuLeuGluIleCysGlyrPheTyrlleGlyAsnCys-GI 238
 Db 809 TTGTTAGGAAAATGTTGGCATTACTTGAGCA-TGCTACTTTCAAGAGGGGAGCTGCAAA 867
 Qy 238 Y-----
 Db 868 AAAAGCATCAAGTGCACCTCCAAATTAAGACTGGAGGACTATCTACCAAGTGTGATGA 927
 Qy 247 uLeuGlyGlyLeuCysGluGluGluIleAsnIleAlaTyMetProProGlyProAspPheAs 267
 Db 928 ATTGGCAAGAAAGTAGGAAAAGGCGCATATTATACATACATCGGAGCCCGATTTTGA 987
 Qy 267 pProThrArgProGluIaThrIleAlaGluAspIleGlyLeuGluGluPheTyrlleGly 287
 Db 988 TCCAACTAAGCAGAACCTCAAGTAGAGAGACATAGGCGTGAAGAGCTTATTAATAA 1047
 Qy 287 eValIleGluIleAspGlyValPheIleGlyValAsnHileuIleArgAlaThrIleAlaLeu 307
 Db 1048 GCGCGCAAGAGATGTGTTCATATCGGAAAGCTCTCTGTAGAGACGCACTGCTCTACT 1107
 Qy 307 uAlaLeuLeuIleGluIleAspSerSerLeuIys---LysValyAspAspIlyGluIysCysG 326
 Db 1108 GGCACCTTTGATTCAGAAATCCAGATCCTTAATTAATTCAGCGCAAGAAAAGAAAATGCA 1167
 Qy 326 uGluIaLeuGluIleuIysSerCysValyAsnProHileGluIleuGluIleuGluIleuCy 346
 Db 1168 AAAAGTCTTGAATAAATAAGTAAAGATTAAAGAAAACATAGAGTTTGGAGATCTTTG 1227
 Qy 346 eValyLeuAsnGlyLeuSerAsnAspGlyThrIlyeIysCysGluIleuGluIleAsnAspI 366
 Db 1228 TAATCAAAATGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1287
 Qy 366 eAsnIlyeThrCysValyIlePheThrSerLeuValIleThrAsnAspArgLeuPheAspPro 386
 Db 1288 AGCAACAGTACTAAATCTTTCTGAAAAATAAGATAAACAATCTC----- 1336
 Qy 386 rLyGluIyAsnAsnGluIleValIleGlyrGluGluIleuProThrPheLeuSerAsnGluIys 406
 Db 1337 -TTCGATCCCGAAGAAACCATTCATGATGTAAGTGTGCGACATTTCTTATGACAGGA 1395
 Qy 406 pCysValIleuLeuGluIleuIysrPheGluIlyeIysCysProAspGlyGluAs 426
 Db 1396 CTGTGCAAGGTTAAGATACAGACTGTTTTATTTGCTCAAGATGAAGACTCTCTTAAAA 1455
 Qy 426 nAlaCysValyAsnIleArgAlaThrCysTyTyLeuArgGlyLeuAspIlaaArgAlaAsnIly 446
 Db 1456 AGAAATGCAAGATGTAAAGCAGCATGTTATTAAGAGAGGCTTATGACCGGCAAAATAA 1515
 Qy 446 eValIleuGluIleuAsnMetArgIlyMetLeuHileGlySerAsnIysSerTrPheGluIly 466
 Db 1516 AGTCTGCAAGAAAATATGCTGGTTATTCAGTGTTCATTAATCAAGTTGGCTTAAGA 1575
 Qy 466 ePheGluGluIleuValIlyeValIyGluIlyeIysCysGluIleuIlyeIysGlySerPh 486
 Db 1576 ATTTCAACAGAAATTAAGTAAAGTATGAGAAACTGAAAGAAAGAAAATAAGAAAGTTT 1635

Qy 486 eSerAsnAspGluLeuPheIleuCysValIleProAlaIysAlaIaArgLeuLeuThr 506
 Db 1636 TTCAACAGATGAAATTAATTTGTCGTGTGTATACAGCCAGAAAAGCAACCCGGTCTTAC 1695
 Qy 506 rHleAspLeuArgMetIysThrIlePheLeuArgGluIleuAspGluIlyAspAspPh 526
 Db 1696 ACATGATCTTCGATGCAACTATCTTTTATTCGCAACAGTTGGATCAAAAGCCAGATTT 1755
 Qy 526 eProThrAspIlyAsnCysValyGluLeuGlyArgIysCysGluIleAspLeuGluIlyAsp-- 545
 Db 1756 TCCGACGTAAGACTGCCAAGGAAATTTAGGAGAAATGCCAAGATTTAGGAAAGCATT 1815
 Qy 546 -SerLyGluIleThrTrpProCysHleThrLeuGluGluIleCysAsnArgLeuGlyTh 565
 Db 1816 CCAAAAAGAAATTAACGTGGCATTCTATACCTGAGACAGAAATGCAATGCTTGGGGAC 1875
 Qy 565 rThrGluIleuLeuArgGluValIleuLeuAspGluHileIysAspThrLeuIlyAspGluI 585
 Db 1876 TACAGAAATTTTGAAGCAGGTTTTTACTTAATGAACACAGAGATCTTTGAAAACCTCAT 1935
 Qy 585 uSerCysValIlyeThrLeuIysGluIysCysAsnIlyeTrsArgArgGlyAspAspAr 605
 Db 1936 AAACGTGTAAACATATTAAAGAAAGATGTAATGAATGCTTAGAAGGGGCGATGACCG 1995
 Qy 605 pPheSerPheValCysValPheGluAsnAlaThrCysGluIleuMetValIlyAspValIly 625
 Db 1996 TTTCTCTTTGTATGTGTGTTTCCAAAACGCTACCTGATAGCTGATGTAATAAGACGTACA 2055
 Qy 625 eAspArgCysGluValaPheIlyAsnIleIysAlaSerTyrlleIleGluPheLeuG 645
 Db 2056 AGATAGGTGCAAAATATTAACAAGAAATATTAAGAAAGTTTCGAAGATTTGTATTTCTTAA 2115
 Qy 645 uAsnAsnThrAsnIlyeIleThrThrLeuGluIlyAsnCysProSerTrpHleThrTyC 665
 Db 2116 AAATTAATACAAATTAACATTAACAACATGAAAGAAATGTCCTCGGCACTAGCTATG 2175
 Qy 665 eAsnArgPheSerProAsnCysProGlyLeuThrIlyeGluIlyAsnSerCysThrIlyeI 685
 Db 2176 CAATAGATTTTCACTTAATTTGTCAGATTTTTCAAAAGAAAGATCTTGTACAAAATTA 2235
 Qy 685 eValyIleArgGluIleuProPheTyTyAspGlyValaLeuGluIleuAspValaIleu 705
 Db 2236 AAACAATTTGAAGCATTATTAATGAAGAAAGCCTTGGAAGACGCTGCAAGTGAAGCT 2295
 Qy 705 uGluGlyIlyeLeuThrAspIlySerIysCysGluProAlaIleuIlyAspArgTyCThrIy 725
 Db 2296 TCGAGGAACCTGATGATGAAGAAACAATGTACTGCAGCATGGAAGGATATTTGTACACT 2355
 Qy 725 IAlaGlyAsnValAsnAsnIleSerIleSerGlyLeuCysIlyeAlaAsnThrIlyAspAs 745
 Db 2356 AGCGGAAATGTAAATAATACCATCATGTCAGAAAGTTTATGCAGAAATTAACACTCAAGT 2415
 Qy 745 nSerGlyIysSerAspGluAspAlaArgIyGluLeuCysGluIlyeValyLeuValyGlu 765
 Db 2416 CAATTAAGAAAGCTGACGAAAAGTGTGAAGAGACTTTGTGAAGAAATTTGATGAAGAA 2475
 Qy 765 IgluGluIleCysValaIaLeuProThrGluIleuGlyGluIleProAlaIleAspLeuIly 785
 Db 2476 AAAAGACCAATGCGACGACATTTACAGCAGAATTAACAACCGGCAACGATCTTAAGAAA 2535
 Qy 785 eAspTyTyLeuThrTyGluIleuIlyeIysValyAspAlaGluIleuIleuIleuIleuIysSe 805
 Db 2536 AGATGTTAAGCATATAGAGAACTTTAAGAAAGAGGCAAAAGAAACATTAAGCAAGTCCAG 2595
 Qy 805 rLeuValIleuSerLeuIleIysIlyAsnGluSerAsnValaSerIysSerAsnSerIysAs 825
 Db 2596 CTTGTTTATTCATCTGTAAAGAAAGCAGAAATTAATACCCGAAATAATATAGCAAAAG 2655
 Qy 825 nIlyAspIlyAsnAlaValaSerAsnGlyLeuGluIleAspThrThrIlyeHileValyIlele 845
 Db 2656 CGAAGATTAAGAAATGCTTCAATGA---AAAGATTAACATTAACATGATGAAGAAATACT 2712

[illegible]

D	61	GAAGAATTCCTTTGGCTTAAATTCGACGAAAAAAATATAATGATATGAATCAAAAAA	120
O	39	LyLeuLyLeuLyTyTCySGInGluLeuThrGluAlaLyLeuAsnIleGluGlnValHis	58
D	121	GAATTTAGAAAAAATATGTGTGTAAGACTTAAACGAGTGCAGAAATTTAAAAACCGAAAAAGTTTAC	180
O	59	ArgLyLeuLyLeuGlyPheCySGInuAspGlyLysAlaAspThrLysCyLysGluLeuLys	78
D	181	AAAAAACTTAAGAGCTTTGTGAAAAATAAAAAGCGAATTCAAAAATGCAAAGACGTGAA	240
O	79	AlaAsnIleGluLyLysCyStnThrIleLysGlyLyLeuLySGluAlaIleLysLys	98
D	241	GA AAAA ACTCACTCAAAAATGTACTCGCAATCAAGSAAAACTTACAGAAAGCATCAAAAAA	300
O	99	LyLeIleGlnIleIleThrAspLysAspCyLysGluAsnGluGlnGlnGlnCyLeuPheLeu	118
D	301	AAAAATTCAGATTTAAACGATGAAAGATTGCAAAAGAGATGAAACAACATCCCTATTTTGTG	360
O	119	GluGluValCySerLysGluLeuLysAspAspCySerAsnThrLeuAsnGluLysCyTyT	138
D	361	GAGGAGACATGCTCCAGCGGAACCTTAAAGATGATTGCAATCTTTGACAAAAATAGCTGAT	420
O	139	GluLyLysLysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeu	158
D	421	CAAAAGAAAGCTGATTAAGTGGCGGAAGAACCTTTTAAGACACATCTCGTGGAGCTCTA	480
O	159	AsnGlySerValIleCySGInuLyLysLeuLySGluIleCySerProValMetGlyArgGlu	178
D	481	ATCAATGAACACTACATGTGAAGGAAAACTCAAGAGGTTTGACTAGCTTGACGACGA	540
O	179	SerAspGluLeuThrAsnLeuCyLysLeuAsnGlnLysGluThrCyLysAsnIleLeuIle	198
D	541	AGTATGATGTTAAACGAAGCTTTGTCTTTATCAAAAAATGACGTGCAAAACATTTGTATTA	600
O	199	GluLyAspLyLysCySGLyThrLeuLysThrAspValSerAlaIleGluLysPhe	218
D	601	GA AAAA CAAAAAATGTAATGCTTAAACAGAGTTTAACGACAGACTTGAC	654
O	219	LyLeLySGlu-----ThrCyLysGluLeuLeuGluGlnCyTyTyPheTyTle	274
D	655	AAGAAAGATGATGATACGAGAAAAATGTTTACACTGCTTGAAACGATCTCTATTTTATAGA	714
O	235	GlyAsnCySGLyAspAspAspIleIleLysCyAlleGlu-----	247
D	715	GGGAATTTGT-----GAAGATATATCAAAATGTATATAATCATCCGAAGCTGTATGAA	768
O	248	-----LeuGlyLyLysCySGInGluAsnIleAlaTyMet	260
D	769	TATTTGCCAGTGTGTATCACTTGGCAGTGAAATGTAAAGAAATAAGATTATTATTATACA	828
O	261	ProProGlyProAspPheAspProThrArgProGluAlaThrIleAlaGluAspIleGly	280
D	829	CATCCGGGATCCGATTTCAATCCAACTAAGTCAAAAGCTACTGTAGCAAGAACATAGGA	888
O	281	LeuGluGluPheTyTyLysLysValGluGluAspGlyValPheIleGlyLysAsnHisLeu	300
D	889	CTGGAAGAGCTTTTAAAAAGCCGCGCAAGAGAGCTTTCATATGTGAAAAGCCCTCTGTA	948
O	301	ArgAspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLyLysLysAsp	320
D	949	AGAGATGACATGCTCTACTGCGCCCTTTGATTCAAAATCTAGATCTTAAAGAGTCAAGCTG	1008
O	321	AspLySGluLyCySGInGluAlaLeuGlnLysSerCyLysAsnProHisGluHisGlu	340
D	1009	GGTAAAGAA---TCGAAAAAAGTTCTTAAAGATACGTAAAGAGTTAAAAAGTCAATGAA	1065
O	341	AlaLeuGluSerLeuCyLysLysAsnGlyLeuSerAsnAspGlyThrLyLysCySGlu	360
D	1066	ATTTTGGAGATTTTGTGATCAAAAATGTAGCTGCTCAAAATGAAATTTGAAAAGTGTAAA	1125
O	361	GluLeuGluAsnAspIleAsnLysThrCyLysIlePheThrSerLysValThrAsnAsn	380
D	1126	GAGTTTAGAGAGGATTTAGCAACAGACTATAAATTTCTTTTGAAAAAATTAAGAAATGAA	1185

QY 381 ArgLeuPheAerProThrLyGlyAAsnAenGluIleValGlyTrpGluGlyLeuProThr 400
 DB 1186 CACCTC-----TCGGATCCGGAGAGACTCATTCATGCTAATGTAACACACA 1233
 QY 401 PheLeuSerAenGluAerPheGlyAlyLeuGluSerTyrcyPheTyrcPheGluLyLeu 420
 DB 1234 TTCTTGTAGTCAATGACTGCACAGGGTTAGTCAGACTGTTTTTATTTAAAAAGTCAA 1293
 QY 421 CyAerPro--AerGlyGluAenAlaCyLeuAenAlaLeuGlyAerGlyAerGlyGly 439
 DB 1294 GCACCTCTTGCACAAAGAA-----TGTAATATCTGAAGGACGATGTTAAAGAGGG 1347
 QY 440 LeuAerAlaAerGlyAenLyLeuValLeuGluGluAenMetAerGlyMetLeuNHISer 459
 DB 1348 CTTGACACACAACTTAATGAAAGCATTCGCAAAAAAGATGTCAGACTGTTTATGCTTCA 1407
 QY 460 AenLySerTrpLeuGluLyAerPheGluGluGluLeuValCyGluLyLeuLyLeu 479
 DB 1408 GGCACAAAGATGTTAAGAACTAGTAAATAATGAAAGAAATGTTCCGAACCTTAA 1467
 QY 480 LyGluGluAenLySerPheSerAerAerGluLeuPheIleLeuCyValGluProAla 499
 DB 1468 ACA-----ACAAAGCATGAGTGTGTTTGTCTATGATGATCCACTT 1509
 QY 500 LyAlyAlaAerGlyLeuThrAlaAerPheAerMetLySerTrpIlePheLeuAerGluGlu 519
 DB 1510 AAGCACTGACAAATACCTTCGAGCTGATATCCAGACAAAGACAGCTTTTTCGGAAACAA 1569
 QY 520 LeuAerGluLyAerGlyAerPheProThrAerLyAenCyLyGluGluGlyAerGlyAer 539
 DB 1570 TTGATCAAAAGCAGACTTCCAAACAGACAAAGATTCGAAGATTAAGAAAGAGT 1629
 QY 540 GlnAerPheGlyGluAerSerLyGluIleThrTrpProCyAerMetLeuGluGluGln 559
 DB 1630 GAACCTTTAGGGAAGATTCAAATCAGATTAAGTGGCCATGCTCACTCAACCAACAG 1689
 QY 560 CyAerAerTrpLeuGlyThrGluIleLeuLySerGluValLeuLeuAerGluNHISerAer 579
 DB 1690 TGTATGCTCTTGGGACCTACGAAATCTTGAAACGGTTTACTAGATGAACACAAAGAT 1749
 QY 580 ThrLeuLyAerGluGluSerCyValLySerTyrcLeuLyGluLyCyAerAlySerTrpSer 599
 DB 1750 ACTTTAAGAACTAGTAAACTGTACGAATATTTAAAGAAATATGTCATTAATGGCT 1809
 QY 600 ArgArgLyAerAerAerPheSerPheValCyValPheGluAenAlaThrCyGluLeu 619
 DB 1810 AGAAGGGGTGATGATCGTTCTCTTTGTATGTGTTACCAAAACGCTACGTTAAGCTG 1869
 QY 620 MetValLyAerPheValAerAerGlyValPheLySerAenIleLyAlaSerTyrc 639
 DB 1870 ATATATAGATGTGTGAAGACAGGTGTGAATTTGAAAAAATATGCCAACCGTCAGAT 1929
 QY 640 IleIleGluPheLeuGluAenAenThrAenLySerTrpIleThrLeuGluAerAerCyAerPro 659
 DB 1930 ATTAAATATCTCTTAATAATTAACAAATTAATAACAAATCAAGATCAAGCAAAATTTGCTCC 1989
 QY 660 SerTrpIleThrTyrcyAerAerPheSerProAenCyProGlyLeuThrLy--Glu 678
 DB 1990 TCATGACCACTCATCTGCGATAGATTTTACCAATTCCTGATCTTAAAGAAAGAA 2049
 QY 679 AenSerCySerTrpLyIleLyLeuNHISerGluProPheTyrcLyAerGlyAlyLeuGlu 698
 DB 2050 ACTTTCTGTCAAAATCTTAATAAATATATGCAACCATTTCAAAAGAGAGGTTTAA 2109
 QY 699 AerAlaLeuLyValGluLeuGluGlyLyAerLeuThrAerLySerTyrcGlyProAla 718
 DB 2110 GATGCTCTTAATAGTAGCTTCAAGGGAATTAAGTATTAATAATATGGAATCTGCA 2169
 QY 719 LeuLyAerGlyTyrcyThrValAlaGlyAenValAenAenAlaSerIleSerGlyLeuCy 738
 DB 2170 TTAGAAGATATTCACAATATTTGAAAAAATGTAAGTATCATCATCAACAGTTATGT 2229

QY 739 LyAlyAenThrLyAerAenSerGlyLySerAerGluAerGluAerGlyAerGlyLeuCy 758
 DB 2230 AAGATATATCCCAAGTAATAAATAAAGCCGATATGAAAGTTAGAAAGAACTTTGT 2289
 QY 759 GlnLyLeuValLySerGluValGluGluGlnCyLeuAlaLeuProThrGluGluGln 778
 DB 2290 CTAAATATAGTGGAAGGGTGGAACAGCAATGTAAATGTTTACCGACGAATGGACAT 2349
 QY 779 ProAlaIleAerPheLySerTyrcLyThrGlyGluLeuLySerAerGlyAerGly 798
 DB 2350 GAGAAAAAGACCTTAAGAATGATTTGAAACATTTGAAAACTTAATAAAGAGCGAG 2409
 QY 799 GlnAlaAerAenLySerSerLeuValLeuSerLeuIleLyLyAerGluSerAerVal 818
 DB 2410 AAAACATATGATTAATCAATCTTTTATCATCTGTTAAGAAAGATGAATAATATACA 2469
 QY 819 SerLySerAerAenSerLyAenLyAerPheAenAlaValSerAenGlyLeuGluAerThr 838
 DB 2470 TCGAAAAATGTAGCAAAAGCAAGATTAAGATACCGTTCCAAACGACCTTCAAGATACC 2529
 QY 839 ThrLyAlyAerGlyIleLeuAerGlyValLyAerPheValSerValThrGluGluGln 858
 DB 2530 ACAGAACTATGAAATAATCTACGAGAGAGATTAAAGATGATATCCATTAACGAATCGAA 2589
 QY 859 AlaLyAlaPheAerPheAlaAlaGluValPheGlyAerGlyAerLyAerGlyAerGly 878
 DB 2590 GCTAAGGCAATTTGATTTGTATGACGAGAGATTTTGAAGATATCTAAGCTTGAAGAAAGA 2649
 QY 879 CyAerAlyLeuGluSerAerPheAerGlyAerGlyAerGlyAerPheLeuGluGluVal 898
 DB 2650 TGTAAATAATGTGAATCAGATTCGACAGATTAAAGAGATTCGAAGATTTAAGAAAGAGTA 2709
 QY 899 CyLyAlyAerGlyAenLyAlyAerAerAenLySerProGluGluValLyProIleGlu 918
 DB 2710 TGTGAAAGATACAGAGATGATGTGCAAAATTAACCACTGAAGTAGAACCCCAAGAA 2769
 QY 919 ThrValThrGluSerThrThrThrThrThrThrThrThrThrThrValAlaAerProLy 938
 DB 2770 ACAGTACAGAAAGCAACAGAGACCAAGCAAGCAACCAAGACGTTATCGATCCGAAG 2829
 QY 939 AlaThrGluCyLySerLeuGlnThrThrAerThrTrpValThrGlnThrSerThrHis 958
 DB 2830 GCAACAGAAATCAATCTTTACAGACAAAGATTCATGATTTACACAGACTTCACACAT 2889
 QY 959 ThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrArg 978
 DB 2890 ACCGACAGCTTACATCATCTTACATCATCAATCAAAAAATTAACATCATCAACAAAG 2949
 QY 979 ArgCyLySerProThrLyCySerThrThrGlyGluGluAerAerAlaGlyAerValLyPro 998
 DB 2950 GCTTGCAAAACCAACCAAGTACAGAGG-----GATGATGACAGAGACGTGAAGCG 3003
 QY 999 SerGluGlyAerGlyMetSerGlyTrpAerValMetArgGlyValIleValAlaMetVal 1018
 DB 3004 AGTAGGATTTGAAGATGAGGGGTGAACCGTATGAGGGGGGTGATATGATGCAATGTT 3063
 QY 1019 IleSerPheMetIle 1023
 DB 3064 ATTTGTTTATGAT 3078
 RESULT 11
 AF033212 AF033212 3080 bp DNA linear PLN 10-SEP-1998
 LOCUS Pnuemocystis carinii f. sp. hominis clone HumsG32 major surface
 DEFINITION glycoprotein (MSG) gene, partial cde.
 ACCESSION AF033212
 VERSION AF033212.1 GI:3560520
 KEYWORDS Pnuemocystis jirovecii
 SOURCE Pnuemocystis jirovecii
 ORGANISM Pnuemocystis jirovecii
 BUKARYOTA; Fungi; Ascomycota; Pnuemocystidomycetes;
 Pnuemocystidaceae; Pnuemocystis.
 REFERENCE 1 (bases 1 to 3080)

AUTHORS Garbe,T.R. and Stringer,J.R.
TITLE Molecular characterization of clustered variants of genes encoding
JOURNAL major surface antigens of human Pneumocystis carinii
MEDLINE Infect. Immun. 62 (8), 3092-3101 (1994)
PUBMED 94314421
7518806
2 (bases 1 to 3080)
AUTHORS Mei,Q., Turner,R.E., Sorial,V., Kilmington,D., Angus,C.W. and
Kovacs,J.A.
TITLE Characterization of major surface glycoprotein genes of human
Pneumocystis carinii and high-level expression of a conserved
region
JOURNAL Infect. Immun. 66 (9), 4268-4273 (1998)
MEDLINE 98380374
PUBMED 9712777
3 (bases 1 to 3080)
REFERENCE Mei,Q., Turner,R., Sorial,V., Kilmington,D., Angus,C.W. and
AUTHORS Kovacs,J.A.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-1997) Critical Care Medicine Dept., National
Institutes of Health, 10 Center Drive, MSC 1662, Bethesda, MD
20892-1662, USA
FEATURES
source Location/Qualifiers
1..3080
/organism="Pneumocystis jirovecii"
/mol_type="Genomic DNA"
/specific_host="Homo sapiens"
/sub_species="hominis"
/db_xref="taxon:42068"
/clone="HUMSG32"
/note="derived from HIV-infected human with P. carinii
pneumonia"
<1..3029
/gene="MSG"
<1..3029
/gene="MSG"
/note="surface antigen"
/codon_start=3
/product="major surface glycoprotein"
/protein_id="AAC34975.1"
/db_xref="GI:3560521"
/translation="ARAYKROAKASGAVYDGEIILALINAGKKNNECKEKELEK
CKTLIDALEKPEKVKLKEPENEKASCKELKEKLTQCTAIKGLTNPATKKNS
DLTDEDCKENQOCLFLEGACPALKDCCNTLRNCKCYQKQDKVAEALIAVAGLI
NETTCBGLKEVCIELSQSDDELTYLCTYQMTCTFYLEKQKCNALAKOVNALAEK
KDELRGKCLPLERCFYRGNCEIDISCKNSCEDCYFLPCDTLAVCEENKIITYH
PGSDNPRTSKSPYVAEDIGLELYKKAEBGVHIGKIPVRATLALINLDPKSO
VGKECEKLNCKELKSHETLDPCNONVAQNIIECKELEKELANSTILPEKIK
NHLISGSEVLPWTLITFLSDNDCTRLSDCFYKSAPLDKECNLKAACVKGLE
AQAENALQKMYGLFYSGKEMFKKLEKIMECESELKTSDELFLCIDPLKVRIL
AADIQARAVFLRKQIDQRDPPTDKCEKELGKCEALGKSDNQLKMPCHTKOCDRL
GTTTELKQVLDENKDTLRHENCTYKLRKCHKNSRGDRFSFVCYQATACGLIV
DDVDKRCVPEKNQASDINSLKNKQIKTSSAANICSMHPYCDRFLPNDLKKG
TECCNLKYYCEPFRKYLDELAKVELOGNSNRKCSALERYTTILKNSDSINS
LCKDNTSKTKTDNEVRKCLKLVVEVQCKMLPDLHEEDKLDDEYTEBKIK
KQAEKTNKSNLVLSFVKDENNTSKNSKDKNTVANGDTHHKILIRGKVD
SVTSEBAKAFDVAEVRGYLDLKERCKLSBCKVCKQKCDLDEGCGKIGVSKLK
PLKVEPHEVLESTTTTTTTTTTVDTPATECKSLQTTDTWITQSTSTSTSTST
TSKITLSTRCKPKCTTGDAEDVKRSEGLKMSG"
3029
/gene="MSG"
/note="putative PCR artifact leads to premature stop
codon"

ORIGIN
Alignment Scores:
Pred. No.: 9,68e-210 Length: 3080
Score: 3396.50 Matches: 654
Percent Similarity: 76.15% Conservative: 141
Best Local Similarity: 62.64% Mismatches: 208
Query Match: 62.77% Indels: 41
DB: 8 Gaps: 12

US-10-654-416-14 (1-1023) x AF033212 (1-3080)
QY 2 ALaar9aIaValIySaTgInaIaIaIaGlyThrGlnaSer-----IleAspGlu 19
Db 3 GCGGCGCGGTCAAGCCGCGAGCTGCAGAAAGCATCAGCGGCTAGTATGATGCTGAA 62
QY 20 HisValIeuAlaIeuIleIeuLygGluAspGlyLeuSerGluGluGluCysGlyLeu 39
Db 63 GAAATCTTTTGGCTTAAATGTCAGGAAAAAATATATATGATATGATGAAAAAAGA 122
QY 40 LeuLySlySTyrcySGInGluLeuThrGluAlaLySLeuAenIleGluGlnValHisary 59
Db 123 TTAAGAAAATTTTAAAGACATTAAACGATCGACAAATTAAAAACGAAAAAGTTCA 182
QY 60 LySLeuLySgLyPheCySGIuAspGlyLySAlaAspThrycYsLySgLyLeuLySAla 79
Db 183 AAACTTAAGAGCTTTGGAAAAATAAAAAGCAGATTCAAAATGCAAGAACTGAAAGA 242
QY 80 AsnIleGluLySlyScyethThrIleLySgLyLySLeuLySgLyAlaIleLySlyS 99
Db 243 AAACCTCACTCAAAATGTACTGCAATCAAGGAAACCTTACAGAGCAATCAAAAAA 302
QY 100 IlegInIleIethrAspLyAspCySleSgLyuAsnGluGlnGInCySLeuPheLeuGlu 119
Db 303 AATTCAGATTTTAACGATGAAGATTGCAAGAGAAATGAACAATGCTTATTTTGGAG 362
QY 120 GLyValCySserLySgLyLeuLyAspAspCySAsnThrLeuAryAsnLyScyTygIn 139
Db 363 GGAACAGATGTCAGCGGGAACCTTAAAGATGATGCAATCTTGAAGAAATAGTGTATCA 422
QY 140 LySlySArGAspLySValIaIaGluGluValLeuLeuAryAlaLeuArySearPheLeu 159
Db 423 AAGAAAGCGTATTAAGCGCGGAAAGCTCTTTTAAAGAGCATGTCGAGGCTTATC 482
QY 160 GlySerValIleCySGIuLySlySleuLySgLyIleCySProValMetGlyAryGluSer 179
Db 483 AATGAACATCATGTGAAGCAAACTCAAGAGAGTTGCAATAGAGTTGATGCAAGAAAG 542
QY 180 AspGluLeuThrAsnLeuCySLeuAsnGluLySgLyThrCySlyAsnIleLeuIleGlu 199
Db 543 GATGAGTTTAAAGAGCTTTGTCTTATCAAAAATAGACGTCAAAACCTTGTATTAAGA 602
QY 200 LySAspLySlyScyGlyThrLeuLyThrAspValSerAlaIaLeuGlySerPheLyS 219
Db 603 AAAAAGAAAAAATGATATGCTCTTAAACAGATGTTAAACGACAGACTTGAG 656
QY 220 LySgLu-----ThrCySleuGluLeuGluGlnGInCySTyrcPheTyriIleGly 235
Db 657 AAGATGAGTTACAGAGAAAAATGTTTACCACTGCTTGAACGATGCTATTTTATAGAGG 716
QY 236 AsnCySgLyAspAspAspIleIleLyScySIIleGlu----- 247
Db 717 AATTGT-----GAAGATATATCAAAATGTAAATATCATCGAAGACTGTATGATAT 770
QY 248 -----LeuGlyGlyLyScySgInGluGlnAenIleAlaTyrcPro 261
Db 771 TTGCGAGTGTGTATACATTGCGAGTGAATAGTAAGAAATTAAGATTAATTAACAT 830
QY 262 ProGlyProAspPheAspProThrArgProGluIaThrIleAlaGluAspIleGlyLeu 281
Db 831 CCGGATCCGATTCATTCACCACTAAGTCAGAAAGCCTACTGTAGCGAGAGACATAGAG 890
QY 282 GluGluPheTyryLySlyValaGluGluAspGlyValPheIleGlyLyAsnHisLeuAry 301
Db 891 GAAGAGCTTTAATAAAGCCGCGAGAGAGAGAGTGTATATGGAAGAGCCTCCTTAAGA 950
QY 302 AspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLySlyAspAsp 321
Db 951 GATGCAACTGCTTACGCGCCTTTTGTATCAAAATCTAGATCTTAAGAGCAAGTGGGT 1010
QY 322 LySgLyLyScyGluGluAlaLeuGluLySerySlySAsnProHisleGluHisleGlu 341
Db 1011 AAAGAA--TGCAGAAAGTTCTTAAGATATAGTAAAGATTAAAAAGTCATGAAATT 1067

Oy	342	LeuGIuSerLeuCyvElybLybAaNgIyLeuSerAaNaPbGlyThrLybCyvEgIuGIu	361
Db	1068	TTGGGAGATTTTGTGTAATCAAAATGTAGCTGCTCAAAATGAAATTGAAATGTAAGATGTAAG	1127
Oy	362	LeuGIaAaNaPbILeaNLyethrCyvAlybIlePheThrSerLybValThrAspAaNaGr	381
Db	1128	TTAGAGAAAGAGTTTACCAACAGTACTAAATTTCTTTTGAAAAAATTAAGATTAACAC	1187
Oy	382	LeuPheAaPProThrLybGlyAaNaNgIyLevalGIyTPGluGIyLeuProThPhe	401
Db	1188	CTC-----TTCGATCCGGAGAGCATTCATCGATGTAATGATTCAGCATTT	1235
Oy	402	LeuSerAaNgIuAaPbCyvAlalybEugIuSerTyTCyPheTyTPheGIuLybCyS	421
Db	1236	CTTAGTGCAATGACTGCACAAGGTTAAGTCAGACTCTTTTATTTAAAGAACCA	1299
Oy	422	Pro---AaPbGlyAaNaAlaCyvElybAaNIleAaGAlaThrCyvTyLybAaGrGIyLeu	440
Db	1296	CCTCTTGCAAAAGAA-----TGTAATATCTGAAGGAGCATGTTATTAAGAGAGGCTT	1349
Oy	441	AaPbAlaAaGrAlAaNLyValIleuGIuGIuAaNaPbAaGrGIyMeLeuNIleGIySerAa	460
Db	1350	GAGCAGACAAGTAATGAGACATTCGAGAAAAAGATGACGACTGTCTTAAGTTCAAGC	1409
Oy	461	LybSerTPbLeuGIuLybPheNgIuGIuLeuValLybValCyvGIuLybElybLyS	480
Db	1410	AAAGATGTGTTTAAGAAACTACTAGAAAAATTAATGAAGAAATGTTCCGAATTTAAAA	1469
Oy	481	GIuAaNLybGIySerPheSerAaNaPbGIuLeuPbIleuCyvValGIuProAlaLyS	500
Db	1470	-----ACAAGCATAGTGTGTTTGTGTAATGATTCACCTTAAA	1511
Oy	501	AlaAlaAaGrLeuLeuThrNIleAaPbLybMeLySThrIlePheLeuAaGrGIuGIuLeu	520
Db	1512	GCAGTCAGAAATACTTGACGCTGATTTCCAAGCAAGACAGCTTTTTCGGAAACAATTG	1571
Oy	521	AaPbGIuLybAaGrAaPbPheProThrAaPbLybAaNCyLybGIuLeuGIyAaGrCyvGIu	540
Db	1572	GATCAAAAGCGAGACTTTCCAAACAGACAAGATTCAGAGAAATTTGGAAGAAAGTGTA	1631
Oy	541	AaPbEugGIyGIuAaPbSerLybEgIuIleThrTProCyvNIleThrLeuGIuGIuNIcyS	560
Db	1632	GCTTATGGGAAGGATTTCAAACTCAGATTAAAGGCGCATGTCAATCCCTAAAACAAGGT	1691
Oy	561	AaNaAaGrLeuGIyThrThrGIuIleLeuLybGIuValIleLeuAaPbGIuNIleLybAaPbThr	580
Db	1692	GATCCCTTTGGGACTACAGAAATCTTGAAACAGGTTTATCTAGATGAACACAAAGACTCT	1751
Oy	581	LeuLybAaPbGrGIuGIuSerCyvAlalySTyTLeuLybGIuLybCyvAaNLySTpSerAaGr	600
Db	1752	TTAAGAACTCATGAAGAACTGTACGAATATTTTAAAGAGAAAATGTCATTAATGCTCTGA	1811
Oy	601	ArgGIyAaPbAaPbAaGrPheSerPheValCyvValPheGIaAaNaAlaThrCyvGIuLeuNec	620
Db	1812	AGGGGTGATGATCGTTCTCTTTGTATGTGTTTCCAAAACGCTACGTAAGCTGATTA	1872
Oy	621	ValLybAaPbValLybAaPbAaGrCyvEgIuValPheLybLybAaNIleLybAlaSerTyTle	640
Db	1872	GTAGATGTATGTAAAGACAGGGTGTAAGATTGAAAAAATGCAACAGCGTCAGATATT	1931
Oy	641	IleGIuPheLeuGIuAaNaAaNLyAaNLyIleThrThrLeuGIuAaGrAaNLyCyvPbSer	660
Db	1932	AATATATCTCTTAABABATTAACAAATTAABAAACAGATACAGCAGCAAAATATTGTCCCTCA	1991
Oy	661	TPbIleThrTyTCyvAaNaAaGrPheSerProAaNCyvProGIyLeuThrLyS---GIuAa	679
Db	1992	TGGCACCACATCTCGATAGATATTTTTACCACATTTCTCGATCTTAAAGAAAGAAAAACT	2051
Oy	680	SerCyvThrLybIleLybLybIleAaGrGIuProPbTyTlybAaGrLybAlaEugIuAaP	699
Db	2052	TTCTGTCAAAATCTTTAAAAAATATTGCGAACCATTTCAAAAGGAAGGTTTAAAGAT	2111

QY	700	AlaIleuValGluLeuEngInGlyLeuThrThraSpLysSerLysCysGluProAlaLeu	719
Db	2112	GCTCTTAAAGAGAGCTTCAAGGAAATTAAAGTATGAAATATGTAATCTGCCTTA	2171
QY	720	LysArgTrpCysThrValAlaGlyAsnValAsnAsnAlaSerIleSerGlyLeuCysLys	739
Db	2172	GAAGAATCTTCCACAAATATTGAAAAAAGTAAAGCATTCATCAATCAACAGTTTATGTAAA	2231
QY	740	AlaIleThrLysAspAsnSerGlyLysSerAspGluAspAlaArgLysGluLeuCysGlu	759
Db	2232	GATAAATCCCAAGATAAATCTAAAAAAGCCGATATAGATTGAAAAAGAGCTTTGCTTA	2291
QY	760	LysIleuValLysGluValGluGluGlnCysLysAlaLeuProThrArgLysGluPro	779
Db	2292	AAATTAGTGAAGAGGTGGAAACGCAATGTAAAAATTGTACCAGCAAGATTGGACATGAG	2351
QY	780	AlaAlaIleAspLeuLysLysAspTyrLysThrThyArgGluLeuLysArgAlaGluGlu	799
Db	2352	GAATAAGACCTTAAAGATGATTTTGAACAATTTGAAAACTTAAAAAAACAGGCGAGAAA	2411
QY	800	AlaMetAsnLysSerSerLeuValLeuSerLeuIleLysLysAsnGluSerAsnValSer	819
Db	2412	ACAAATGATTAATCAATCTGTTTTCATCTTCCTTAAAGAAAGATGAAATTAATACATCG	2471
QY	820	LysSerIleAsnSerLysAsnLysAspLysAsnAlaValSerAsnGlyLeuGluAspThrThr	839
Db	2472	AAAAATGATGCAAAAGACAGGATTAAGAAATCCCTTTCAACGAGACTTCAAGATACACA	2531
QY	840	LysIleValLysIleLeuArgArgGlyValLysAspValSerValThrGluLeuGluAla	859
Db	2532	GAACTATGAAATATCTACGGAAGGAGTTTAAAGATGTAATCCGTACAGAACTCGAAAGCT	2591
QY	860	LysAlaIleAspLeuAlaAlaGluValPheGlyArgTyrValAspLeuLysGluArgCys	879
Db	2592	AAGCATTTGATTTGTGAGACAGAGATTTTGAAGATATCTAGACTGTAAGAAAGAACTGT	2651
QY	880	AsnLysLeuGluSerAspCysArgIleLysGluAspCysLysAspLeuGluGluValCys	899
Db	2652	AATTAATTTGGAAATCAGATTGACAGATTGAGAGGATTGCAAGATTTAAGAAAGAGTATGT	2711
QY	900	LysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluValLysProIleAspGluThr	919
Db	2712	GGAAGATACAGAGATGATGTTCCAAATTAAACCACTGAAGATGAAGCCGACAGAAACA	2771
QY	920	ValThrGluSerThrThrThrThrThrThrThrThrValAlaAspProLysAla	939
Db	2772	GTGACGAAGAACACACAGACGACACGACGACCAACACGACCGTTACGATCCGAAGGCA	2831
QY	940	ThrGluCysLysSerLeuGlnThrThrAspThrTyrValThrGlnThrSerThrHisThr	959
Db	2832	ACAAATGCAATCTTTACAGACAAACAGATTCAGATTCACAGACTTCGCACACATACC	2891
QY	960	SerThrSerThrIleThrSerThrIleThrSerLysIleThrLeuThrSerThrArgArg	979
Db	2892	AGCAGCTTACCATCACATCTTACACATACATCAAAAAATTAACCTCACATCAACAGGCGT	2951
QY	980	CysLysProThrLysCysThrThrGlyGluGluAspAspAlaGlyAspValLysProSer	999
Db	2952	TGCACCAACCAACCACTGTACACAGCG-----GATGATGACAGAGACGTGAACCCGAGT	3005
QY	1000	GluGlyLeuArgMetSerGlyTyrAsnValMetArgGlyValIleValAlaMetValIle	1019
Db	3006	GAGGAGATTGAAGATGAGTGGGTGAACGTCGTGAGGGGGGTGATGATGCAATGTTATT	3065
QY	1020	SerPheMetIle 1023	
Db	3066	TCGTTCATGATT 3077	
RESULT 12			
AR438434			
LOCUS AR438434 3042 bp DNA linear PAT 20-FEB-2004			
DEFINITION Sequence 1 from patent US 6664053.			
ACCESSION AR438434			

```

VERSION      AR438434.1  GI:42663289
KEYWORDS
SOURCE       Unknown.
ORGANISM     Unknown.
REFERENCE    1 (bases 1 to 3042)
AUTHORS     Kovacs, J.A., Huang, S., Masur, H., Fischer, S.H., Gyll, V.J. and Mel, O.
TITLE        Identification of a region of the major surface glycoprotein (MSG)
JOURNAL      Patent: US 6664053-A 1 16-DEC-2003;
FEATURES     Location/Qualifiers
             source          1..3042
                        /organism="unknown"
                        /mol_type="genomic DNA"

Alignment Scores:
Pred. No.:      7,476-169      Length:      3042
Score:          2762.50        Matches:     541
Percent Similarity: 69.10%     Conservative: 177
Best Local Similarity: 52.07%   Mismatches: 280
Query Match:    51.05%        Indels:      41
DB:             Gaps:         18

US-10-654-416-14 (1-1023) x AR438434 (1-3042)

OY      1 MetAlaATgAlaValLysArgGlnAlaAlaGlyThrGlnAsnSerIleAspGluGlnHis 20
DB      1 GTGGCGCGGCGCGGCTTAAGCGGCAAGGTAACAGAGACA--TCAGAGATGATGAGAGAGAA 57
OY      21 Val-----LeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGlnGluCysLysLys 38
DB      58 GTGCGCTCTTTTGGCTTTAATCTAAAGAAAGATTTAAAGATGAATAAAGCAAGAA 117
OY      39 LysLeuLysLysThrCysGlnGlnLeuThrGlnAlaLysLeuAsnIleGluGlnValHis 58
DB      118 AATATTGAAAAACATTCGCAAGATTTAAGTGAAGCAATCTTAACCCAGAACAACTACAT 177
OY      59 ArgLysLeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLys 78
DB      178 GAAAGATTAAAGATTTCTGTATGCAAAAACGATTAATAAATGTAAAGAACTAAAA 237
OY      79 AlaAsnIleGluLysLysCysThrIleLysGlyLysLeuGluAlaIleLysLys 98
DB      228 AAAAAATGTTGAAAAAAATGCGGTATTTTAAACAGAAATTGAAGAAATGGGTGAAAAAG 297
OY      99 LysIleGlnIleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeuPheLeu 118
DB      238 GAAGCTTCAAATTGAAAAATGATGAGTACAAAAAATGAACAACAGTCTTGTTTTA 357
OY      119 GluLysValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyr 138
DB      358 GAAGAAGCATGCTCT--GATCTTACAAAGAAATTCACACGATTTTAAAGAAACCAATGTAT 414
OY      139 GlnLysLysArgAspLysValAlaGluGlnValIleLeuArgAlaLeuArg----- 155
DB      415 CAGGAATTAAGCGGATTAAGTACCAAGAAAGGTTCTTTTAAAGATATATATAAAGAAAAAGAT 474
OY      156 ---SerAspLeuAsnGlySerValIleCysGlnLysLysLeuLysGluIleCysProVal 174
DB      475 TTTAAAGATTAATAATTC-----TGTGAATAATTAACGTGAAGATATACGTCAAGAA 525
OY      175 MetGlyArgGluSerAspGluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLys 194
DB      526 TTAAGTCAAAATGAGACGAATTTGATGAATATATGTTTGAATCAAAAAAATACGTGTAT 585
OY      195 AsnIleLeuIleGluLysAspLysLysCysGlyThr-----LeuLysThrAspVal 211
DB      586 AATCTTGAAGAAACGCAAAAGTGTGAATCTTTCAAAATCTTTAAACGGAATTT 645
OY      212 SerAlaAlaLeuGlySerPheLys--LysGluThrCysLeuGluIleLeuGlnGlnCys 230
DB      646 AAAACAAATTAAGAAAGATGAACAACATAAATAAATGCCCCATTATATATGAAGAAATGC 705

```

```

OY      231 TyrPheTyrIleGlyAsnCysGlyAspAspAspIleIleLysCysIleGluLeuGlyCly 250
DB      706 ATTTTTATATGATGAAGATGTGGAAACGATTCA--CTGAAGTGTAGTAATGGAAAAA 762
OY      251 LysCysGlnGluGlnAsnIleAlaTyrMetProProGlyProAspPheAspProThrArg 270
DB      763 AATGTGCAAGGAAAAATATTATCTTACACATTAATCAATTAAGAGTTTGATCTTATAGA 822
OY      271 ProGluAlaThrIleAlaGluAspIleGlyLeuGluGluPheTyrLysLysValGluGlu 290
DB      823 CCAGAAATTAATTAAGCAAGAAAGTACCTTAAGAGAAATTTATAGAAAGCAGCAAGA 882
OY      291 AspGlyValPheIleGlyLysAsnHisLeuArgAspAlaThrAlaLeuLeuAlaLeuLeu 310
DB      883 GAAGAACTCTTGTGGGAAACCTTTACACAGCAATGCTACCTGCTTGGTGCATTTTGG 942
OY      311 IleGlnAspSerSerLeu--LysLysLysAspAspLysGluLysCysGluGluAlaLeu 329
DB      943 ATTCAAGATCCATCTTTCAACTCAACGACTCAACACTCAACAAAGAAAAATGTAATAATTTCTT 1002
OY      330 GlnLysSerCysLysAsnProHisGluHisGluAlaLeuGluSerLeuCysLysLysAsn 349
DB      1003 GAAGATAAATGTAAATAATTTAAAGAACATGATATTTAAAGAGTCTATGCGAGATTAAT 1062
OY      350 GlyLeuSerAsnAspGlyThrLysLysCysGluGluLeuGlnAsnAspIleAsnLysThr 369
DB      1063 AATCAAAATTAAGATTAAGCAACAAAAATGCGAAGAACTTACTACAGATTTGAAGAAACA 1122
OY      370 CysLysIlePheThrSerSerValThrAsnAsnArgLeuPheAspProThrLysGlyAsn 389
DB      1123 GTAAATTTTTCATTCAAAAACCTTAATGATTAATTTTGGCGATGCAATTAATAAT 1182
OY      390 AsnGluIleValGlyTyrGluGlyLeuProThrPheLeuSerAsnGluAspCysAlaLys 409
DB      1183 GATGAATATTAAATGCGGGAATTTTCAACGTTTCTTAACCAATTAAGATTTGTACAAA 1242
OY      410 LeuGluSerTyrCysPheTyrPheGluLysLysCysProAspGlyGluAsnAlaCysLys 429
DB      1243 TTAGATCTGATGCTTATTTTAAAAAAAGCGT--AGAACGAAATCGCATGCAAG 1299
OY      430 AsnIleArgAlaThrCysTyrLysArgGlyLeuAspAlaAspAlaAsnValLeuGln 449
DB      1300 AATATCAGAGCATGCTGTAAGAGAGACTTACACATTAAGCAATTAAGATTTACAA 1359
OY      450 GluAsnMetArgLysMetLeuHisGlySerAsnLysSerTyrLeuGluLysPheGlnGln 469
DB      1360 AAAGAAATGCGAGGAATGCTGCATGTTCAATAAATACATGCGTTAGTGTCCAAAAA 1419
OY      470 GluLeuValLysValCysGluLysLeuLysLysGluLeuLysGlySerPheSerAsnAsp 489
DB      1420 AAATCATAGAAAGTGTGCAAAAAAGTGAAGAAAGATTAAGAGATTTTCCGAGTAT 1479
OY      490 GluLeuPheIleLeuCysValGlnProAlaLysAlaIleAspLeuThrHisAspLeu 509
DB      1480 GAATTAATTTGCTTATGTTGTAACCAACATCAAGAGCATGCTTCAATTTCCGATGATCTT 1539
OY      510 ArgMetLysThrIlePheLeuArgGlnGlnLeuAspGlnLysArgAspPheProThrAsp 529
DB      1540 CGGATGAATAATCTTTTTCAGAGACGATTTGAACAGAAAGCGAGATTTTCCAGTGA 1599
OY      530 LysAsnCysLysGluLeuGlyArgLysCysGluAsnLeuGlyGluAspSerLysGluIle 549
DB      1600 GAAGCTGCGGAAGATTAATTAAGAAATGTCAGCTCTTAAGAAAGGATTTCAAAAAATTT 1659
OY      550 ThrTyrProCysHisThrLeuGluGlnGlnCysAsnAspGluGlyThrThiGluIleLeu 569
DB      1660 GAATGCGCATGTCTATCACTTAAGCCAAATTTGTATGATTAAGAAACCGCTTAAGAAATTTG 1719
OY      570 LysGlnValLeuLeuAspGluHisLysAspThrLeuLysAspGlnGlnLysCysValLys 589
DB      1720 AAAGAACTTTACTAATTAAGAACTATAGATATTTGAAAAATCAAGCAATTTGTGCAATG 1779

```

Qy 590 TyrluYegluYsCyAsnlyeTrpSerArgArglyAspAArgPheSerPheVal 609
 Db 1780 TATTTGAAGGAGAAATGCAATGATGGCTTAGAAGAGAAATGAACGTTCTCTCTTTA 1839
 Qy 610 CyValAlphegluAenAlaThrCyeglulueMetVallyAspVallyAspArgCyeglul 629
 Db 1840 TGTGCTTTGGCAAAATAGACACTTGCAGAAATATGCTAGAAATGCAATGCAAA 1899
 Qy 630 ValPheylAsnAllyeValSerTyrlellegluPheleuGluAenAlaThrAsn 649
 Db 1900 ATATTGTAAGAAACATTAAAAATACCAAGCATTTGAT-----AGTAAACT 1947
 Qy 650 LyelelThrThrleuGluArgAsnCyProSerTrpHisThrTyrCyAsnArgPheSer 669
 Db 1948 AAAATAGAAATATAGGACATATGTCCTATTGGACCCACCTGCCATAGATTGGGA 2007
 Qy 670 ProAsnCyProglYleuThrLyeglulueAsnSerCyThrlylylelyeVHleArgGlu 689
 Db 2008 CCCAATTGCCCGGACTTTGAAAAA---ATTAATGTGAAGACTTTGAAAAATATTGGCAA 2064
 Qy 690 ProPheTyrlyAsArglyAlaLeuGluAspAlaLeuysValGluLeuGlnlyLyleu 709
 Db 2065 CCTATTATTAGCAAGAACCTTGAAAAAGCACTTATATTGTTAGGTTAGGACATCTT 2124
 Qy 710 ThrAspLySerLyCyeglulProAlaLeuLyAsArgTyrCyThrValAlaGlyAsnVal 729
 Db 2125 GATTAAGAAAAAACTGCCAAAAACAATCTTATAGTACTTACACTATGGGATCAAAACA 2184
 Qy 730 AsnAsnAlaSerleSerGlyLeuCyAllyAlaAnthrlyAspAsnSerGlyLySer 749
 Db 2185 GGAATTAACACTTAAGGTTTTTGTGAAGCTTACTGATTAACAAT----- 2232
 Qy 750 AspGluAspAlaArglyBgluLeuCyeglulysleVallyBgluValGluGlnCys 769
 Db 2233 ---GAAACATTTAAGATTAACCTTTGGCAAAAGCTTACTGATTAACAAT----- 2289
 Qy 770 LysAlaLeuProTrpGluLeuGlyGlnProAlaAlaAspLeuLylyAspTyrLyThr 789
 Db 2290 CAAGGATTTATCAAAAGAACTTGAAGAAAAAGCAAAATGATTTTGAAGAAAAACATTAAGAT 2349
 Qy 790 TyrGluGluLeuLylyAsArglyAlaGluAlaMetAsnlySerSerleuValleuSer 809
 Db 2350 TATGAAAAAGTAAAGAAAGATACAAAAATGCAATGAAAGAAACAATCTCGTTTTTCA 2409
 Qy 810 LeuilelylyAsnGluSerAsnValSerlySerAsnSerlyAsnlyAspLyAsn 829
 Db 2410 ACAACTAAATCAACAGATTAATTAACAGAAAAAGCA-----GTCAAGCCT 2454
 Qy 830 AlaValSerAsnGlyLeuGlnAspThrThrlyshVallySileleuArgGlyVal 849
 Db 2455 AGTACGCTTACTGTATTCAGATTAATTTGTA---CATTTTAACTTTGTAAGAAATGCA 2511
 Qy 850 LysAspValSerValThrGluLeuGlnAlaLyAlaPheAspLeuAlaGluValPhe 869
 Db 2512 AAA---GTTCAAGTGACAGAAAGAAAGCAAAAGGTTGATTGGTAGCACTAGCATTC 2568
 Qy 870 GlyArgTyrValAspLeuLyeglulArgCyAsnlyleuGluSerAspCyArgIlelye 889
 Db 2569 AGCTTTATATGAGAGTTAAAGAAACGTGTACCACTTAAAGATGATTCGAATTGGA 2628
 Qy 890 GluAspCylyAspAspLeuGluGluValCyelyslylyleuAllyValaCyArgAsnleu 909
 Db 2629 AAAAGAGTAA---TGTAAAGACAGTGCAAGAGATGAAGAAATATGTTTAAATA 2685
 Qy 910 LysPheleuGluVallyPheProHleGluThrValThrGluSerThrThrThrThr 929
 Db 2686 GAACACAGTAAAGTAAAGCACAATGAATTAACACAGTAAGGAAACCAACATTAACAACA 2745
 Qy 930 ThrThrThrThrValAlaAspProlyAlaThr-----GluCyelysSer 944
 Db 2746 GTCAACAGAAACGTCAAGAAAGAAAGAAACAGTAAGAGACGAGAGAAATGCAAACTT 2805
 Qy 945 LeuGlnThrThrAspThrTrpValThrGlnThrSerThrHisThrSerThrSerThrIle 964

Db 2806 CTCAGCACAACAGACAGTGGGTCAACAAGACGTCACCATACACACACTCCACACT 2865
 Qy 965 ThrSerThrIleThrSerlyleThrleuThrSerThrArgArgCylyAspProThrLy 984
 Db 2866 ACGTCCACAGTACGTAAAGATTAACCTACCTCGACGAGCGGCTTAAGCTTACAGAG 2925
 Qy 985 CyethrThrGluGluGluAspAlaGlyAspVallyAspProSerGluGluLyAsArgMet 1004
 Db 2926 TGTACGACAGAGAGAGAAATGAACGAGAGAGTGAAGCGAGTGAAGCGGCTGAGATG 2985
 Qy 1005 SerGlyTyrAsnValMetArglyValIleValAlaMetVallyleSerPheMetIle 1023
 Db 2986 AGTGGTGAGAGTGAAGAGAGGCGGTGTATTAGCATGATGATTCATTCATGATAT 3042
 RESULT 13
 AF038556
 LOCUS
 DEFINITION
 PNEUMOCYTIS CARINII F. SP. HUMANIS VARIANT REGIONS OF MAJOR SURFACE GLYCOPROTEINS (MSG1, MSG3, MSG4) GENES, PARTIAL CDS.
 AF038556.1 GI:3560524
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 PNEUMOCYTIS JIROVECI
 PNEUMOCYTIS JIROVECI
 Eukaryota; Fungi; Ascomycota; Pneumocystis; Pneumocystis.
 REFERENCE
 AUTHORS
 TITLE
 1 (bases 1 to 12792)
 Mel, O., Turner, R.E., Social, V., Kilmington, D., Angus, C.W. and Kovacs, J.A.
 Characterization of major surface glycoprotein genes of human Pneumocystis carinii and high-level expression of a conserved region
 JOURNAL
 Infect. Immun. 66 (9), 4268-4273 (1998)
 MEDLINE
 96380374
 PUBMED
 9712777
 REFERENCE
 2 (bases 1 to 12792)
 Mel, O., Turner, R., Social, V., Kilmington, D., Angus, C.W. and Kovacs, J.A.
 Direct Submission
 Submitted (12-DEC-1997) CCMD, NIH, Building 10, Room 7D43, MSC 1662, Bethesda, MD 20892-1662, USA
 location/Qualifiers
 1. 12792
 /organism="Pneumocystis jirovecii"
 /mol_type="genomic DNA"
 /specific_host="Homo sapiens"
 /sub_species="hominis"
 /db_xref="taxon:42068"
 <668. 3712
 /gene="msg1"
 /note="multicopy gene family is expressed by gene rearrangement"
 <668. 3712
 /gene="msg2"
 /note="glycoprotein A; expressed from a single putative leader sequence located upstream of coding region; variant region is in frame with the leader sequence"
 /codon_start=1
 /product="variant region of major surface glycoprotein"
 /protein_id="AAC34980.1"
 /db_xref="GI:3560525"
 /translation="VARAAYROYTGAAGSUNDEEYRLALILKENSXDKKCEKLEKH CKEISEANLTPBOVHEKLDPCDSKRDKCKELKRVYKCGDPFTELEBYLVKENS NLKNDCEKNEQCLFLEACSDLTNXCNDLRNKCYONKDKVAKEVLNLIKNNFK DKNSCNKLENYCOELSQMSDELMKLCPDQNKTCNDLVKSTQKESFKNLKTBIKT KEDEOLKKKRPILYEBCTFYDESQNSNLRKSELEKCKQKNITVTSYSGFPIEBE ITTAAEYDLGIIYKKAEBEGLVGGPLPAPATLVAVLIQDPSLTITSTYKKECKKIL EDCKNKKEHDIILGICEDYANAKDKDKKEELSTIECKPISITLMIHFGDGN KNDGITKMGUNSTFLSKDCTKYLETYLYREKCSBSTAKPNIRACYKXGDLTLANE VLOKEMTGAGHNSKNTWLSGFQKLELVCKKVKKENGVPSNELFLVCVQPSGABU LSHLRMKTTFLDLDLNRKDPVYKDECEBELKKCEALRDSKKEIWPCHTSLGNCQD LRNAKEIKELLNENKDIILKNOECGMYLREKCNENWERNRERNRFSLICALONRYCRIM

KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3006)
AUTHORS Kovacs, J. A., Huang, S., Masur, H., Fischer, S. H., Gill, V. J. and Mei, Q.
TITLE Identification of a region of the major surface glycoprotein (MSG)
JOURNAL Patent: US 6664053-A 3 16-DEC-2003;
FEATURES location/Qualifiers
source 1. 3006
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.:	9.1e-146	Length:	3006
Score:	2404.50	Matches:	479
Percent Similarity:	65.37%	Conservative:	195
Best Local Similarity:	46.46%	Mismatches:	320
Query Match:	44.44%	Indels:	37
DB:	6	Gaps:	19

US-10-654-416-14 (1-1023) x AR438435 (1-3006)

OY 1 MetAlaArGaLaVaAlLyArGGLnAlaAglYThrGlnAsnSerTLeApGluGluHis 20
DB 1 GTGGCGCGGCGGTCAACGCGCGGCTGCA--GCACAGAAATGAGTTGAAGAAATAT 57
OY 21 ValLeuAlaLeuAlLeuLeuGluAspGlyLeuSerGluGlnGluCysAlaLeuLeu 40
DB 58 CTTTGGCTTGTGATTTAAAGAAATGAGTAAGAAATTAATGATTAAGTAAAGAAAGTTG 117
OY 41 LysLysYrCYSGInGluLeuThrGluAlaLysLeuAsnLeuGluGlnValHisArgLys 60
DB 118 AAAGAGTATTGTGAAGATTAAAGAAATGTAACAAAGAAACCAAGAAAGCTGAAGAAAG 177
OY 61 LeuLysGlyPheCYSGInAspGlyLysAla--AspThrLysCYsLysGluLeuLysAla 79
DB 178 TTAGACGGAATCTGCAAGATGATTAACAATAGAACAAATGCAAGAAATCAGAAACA 237
OY 80 AsnLeuGluLysCYsThrThrLLeuGlyLysLeuLysGluAlaLeuLeuLysLys 99
DB 238 AAGGTTAAAGCAAGATGTAAGTTTTCACACAGAACTTGATTAAGCAAGTCAAGAAAGGA 297
OY 100 IleGlnIleIleThrAspLysAspCYsLysGluAsnGluGlnGlnCysLeuPheLeuGlu 119
DB 298 GCTTCAACATTGAAGATPATGATTTGTAAGAAAGATGAACGACATGCTGTTTGGAG 357
OY 120 GlyValCYsSerLysGluLeuLysAspAspCYsAsnThrLeuAspAsnLysCYsTyrGln 139
DB 358 GGAGCATGTCCAAACGAACCTTAAGATTAATGAACTGACGAGAAATTAATGTTATCAA 417
OY 140 LysLysAspAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsn 159
DB 418 AAAAAACAGACGACGACGAGAAAAAGCTTTTAAGGTAAGTGAAGGGAACCTTAAG 477
OY 160 GlySerValIleCYsGluLysLeuLysGluIleCYsProValIleGlyArgGluSer 179
DB 478 GATAAAACACATGCAAAATAAGTTAAAGGGGTTTGTCAAGATTCACAAAGAAAGT 537
OY 180 AspGluLeuThrAsnLeuCYsLeuAsnGlnLysGluThrCYsLysAsnIleLeuIleGlu 199
DB 538 GATGAGCTAAATTAATTTATGCTTGAACAGAAAAAAGCTGTGAGATCTGTATCTAAG 597
OY 200 LysAspLysLysCYsGlyThrLeuLysThrAspValSerAlaLeuGly-----Ser 217
DB 598 AAAAGATTCAAATGCAAACTCTCAAGAGAAAGAAATGATCTAGCTTGAAGAAAGAAAT 657
OY 218 PheLysLysGluThrCYsLeuGluLeuLeuGluGlnGlnCYsTyrPheTyrIleGlyAsnCYs 237
DB 658 TTATTAAAAAGAAAAATGTTTATTATTCTTGAAGAAATGTTACTTTTAAGGTCAAAGTGT 717

OY 238 GlyAspAspAspIleIleLysCYsIleGluLeuGlyLysCYsGlnGluAsnIle 257
DB 718 ---GAAACAGATCAGCCAAAGTTGTAAGCTTGCAGAACCAATGTCAAGAAATCTC 774
OY 258 AlaTyrMetProProGlyProAspPheAspProThrArgProGluAlaThrIleAlaGlu 277
DB 775 GTTATTCAGACACAGGTTCACTTGTATCTTCTCGAAATTAAGATTAAGCTTGAACGAA 834
OY 278 AspIleGluLeuGluGluPheTyrLysLysValGluGluAspGlyValPheIleGlyLys 297
DB 835 GAAATAGACCTAGAAAAATGTATGATGAGACGATGAAAAAGGAATTCATATTGGAAG 894
OY 298 AsnIleLeuArgAspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerLeuLys 317
DB 895 CCATCAATAAAAAGTAGAAGTCGCTTTATGTGCATTTATTAACCAAGATGATGCT----- 948
OY 318 LysLysAspAspLysGluLysCYsGlyGluAlaLeuGlnLysSerCYsLysAspProHis 337
DB 949 CAAATATCTTTTAAGATCAATGTGAAGATCTTATTAAGAAAAAATGTGAAACTTTTAA 1008
OY 338 GluHisGluAlaLeuGluSerLeuCYsLysAsnGlnLysSerAsnAspGlyThrLys 357
DB 1009 GAGCATATTATTAAAAAGATTTATGTAGTAAGACTATCAGCATTAATCCAAAGAA 1068
OY 358 LysCYsGluGluLeuGlnAsnAspIleAsnLysThrCYsLysIlePheThrSerLysVal 377
DB 1069 AAATGCCAAGAACTTAATAGAGCTTA-----ACAACCCGTATTTTAAGCTTTCTTAA 1122
OY 378 ThrAsnAsnArgLeuPheAspProThrLysGlyAsnAsnGlnIleValGlyTyrGluGly 397
DB 1123 AGATTTGAAGAAATTTTCGCTCCAGCT---AATGTAAGGAAATTTATGTTGGCATATG 1179
OY 398 LeuProThrPheLeuSerAsnGluAspCYsAlaLysLeuGluSerTyrCYsPheTyrPhe 417
DB 1180 TTGCATACATTTCTTGTTGAAGAGAGGTGACGAAAGCTGTGCGATGTTGTTTATTTG 1239
OY 418 GluLysLysCYsProAspGlyGluAsnAlaCYsLysAsnIleArgAlaThrCYsTyrLys 437
DB 1240 AAAAGCCAAAGCTCA---CTTGAAGAGCCCTGCAATTAACCTTAAGCAGCATGTTATAA 1296
OY 438 ArgGlyLeuAspAlaArgAlaAsnLysValLeuGlnGluAsnMetArgGlyMetLeuHis 457
DB 1297 AAAGGCTTGAAGCAGTAGCAATGAACATTAAGATTAAGTGAAGGGAAGAAATTCGAA 1356
OY 458 GlySerAsnLysSerTyrPheGluLysPheGlnGlnGluValLysValCYsGluLys 477
DB 1357 GGTTCAAATAGAACATGCTGTAACCTTCACAAAAAAGCTGTGTAAGTGTGTAAG 1416
OY 478 LeuLysLysGluAsnLysGlySerPheSerAsnAspGluLeuPheIleLeuCYsValGln 497
DB 1417 ACGAAAGGAGAA-----AGTATGAAATTAATTTGATCTATGATTAAGAAC 1458
OY 498 ProAlaLysAlaAlaArgLeuLeuThrHisAspLeuArgMetLysThrIlePheLeuArg 517
DB 1459 CCAATTAAGAAAGGCTCTTACAGTGTCAACAGATTTGCAATGAGAGGCACTTGCTTGCAG 1518
OY 518 GlnGlnLeuAspGlnLysArgAspPheProThrAspLysAsnCYsLysGluLeuGlyArg 537
DB 1519 GAGCATTTGAACGAAAAACGAGATTTTCCACAGAAAAAGATGTAAGAAATTAAGAGAA 1578
OY 538 LysCYsGlnAspLeuGlyGluAspSerLysLysIleThrTyrProCYsHisIleThrLeuGlu 557
DB 1579 AAATGTAGGCTTGAAGAAAGATTCAGAGAAATTAATGATGCTATGTAATACGTTAA 1638
OY 558 GlnGlnCYsAsnArgLeuGlyThrThrGluLysLeuLysGlnValLeuLeuAspGluHis 577
DB 1639 CAGCATTCGATTCGGCTGAAGACATAGACACTTAAGAGAGAGATTCCTAAAAAGAAAT 1698
OY 578 LysAspThrLeuLysAspGlnLysSerCYsValLysTyrLeuLysGluLysCYsAsnLys 597
DB 1699 AAAGGATTTTAAAGATGAAGAAATAGCTGCAAGAGAAAGAGTGAAGAAAGATGTGAA 1758
OY 598 TrpSerArgArgGlyAspAspArgPheSerPheValCYsValPheGlnAsnAlaThrCYs 617

```

Db      1759 TGGTTTGAAGAGAAAATAAATTTTTCGGCTTGCTTCGACTTGGAAGCTTGTTC 1818
Qy      618 GlnuemetValyluAspValyluAspArgCysGluValPheyluValuAsnIleluAla 637
Db      1819 AAAAAAGATCACTAATAAGTTGAATCTTAATGTAATATGTAAGGACAATGAAACT 1878
Qy      638 SerTyrIleIleGluPheleuGluAsnAsnThrAsnIleThrThrleuGluArgAsn 657
Db      1879 ATGAACCTTATTAAGTAAATAGCTAAAAAAGAGAAAAAATh----- 1920
Qy      658 CysProSerTyrPheIleThrTyrCysAsnArgPheSerProAsnGlyProGlyLeuThr-- 676
Db      1921 TGTGAATTTTGGGCTCCATATTGTAAAAAGTACAGACAAATTTGAAAAAATCTTAAAAAC 1980
Qy      677 -----LysGluAsnSerCysThrIlelelylelyshAsnGluProPheTyrIleArg 694
Db      1981 GGAGGAAAAAGATGGGCAATGCAAAAACTCAATAAAAAGTCAAAATCATTCCTTGAAGAA 2040
Qy      695 LysAlaIleuGluAspAlaIleuLysValGluIleuGluGluIleuThrAspLysSerIys 714
Db      2041 GAAGCTTATGAAAAATAAGTTGTAGAAAGATTGAAAGGTAATTATCAACGTAAGAGAA 2100
Qy      715 CysGluProAlaIleuLysArgTyrCysThrValAlaGluAsnValAsnAsnAlaSerIle 734
Db      2101 TGTAAACAATACCTTAATATATATACGTACCAATGAAAAAGCAGAGAAAT--GGGTTG 2157
Qy      735 SerGlyLeuLysCysValuAlaAsnThrLysAspAsnSerGlyLysSerAspGluAspAlaArg 754
Db      2158 GAAACTTTGTGCAAA-----AGCAAAAGAAAAACAC--AAGAGTGCATTAAAGTTAGA 2208
Qy      755 LysGluLeuLysCysGluLysLeuValyluGluGluGluIleuLysValaIleuProThr 774
Db      2209 GAAGAACTCTGTGAAAAAGCTTAATAACGTTAAAAAGAAAAATGCTCAAAATTGAAGAC 2268
Qy      775 GluIleuGluIleuProAlaAlaAspLeuLysLysAspTyrIleThrTyrGluGluLeuLys 794
Db      2269 GAGCTTGAAGAAATGAAAGAGCTCTTAGAAAAAGAAAGAAAGTATTAATAAATAATTA 2328
Qy      795 LysArgAlaGluGluAlaMetAsnLysSerIleuValIleuSerLeuIlelyluAsn 814
Db      2329 GAAGAAACAGAAAAAGCCATGGAAGATGCCAAACCTTATTTATCGAGAGCGAAAGGACT 2388
Qy      815 GluSerAsnValSerLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGly 834
Db      2389 GATATATATATATATATAG-----TCAGTAAATTAAGACTCATCTGATACCTTAAGGA 2442
Qy      835 LeuGluAspThrThrLysIleValyluIleuLysArgGlyValyluAspValSerVal 854
Db      2443 GGAAAAAGGCACAAAC---GCAATTAACTGTAAGAAAGAAATGCAGAA---GTGCATGTA 2496
Qy      855 ThrGluIleuGluAlaLysAlaPheAspLeuAlaIleGluValPheGlyArgTyrValAsp 874
Db      2497 ACAGAAAAAGAAATTAAGCAGCACTTTGATTGTTAGACAGAGCATTTGATCTCTATCTGA 2556
Qy      875 LeuLysGluArgCysAsnLysLeuGluSerAspCysArgIleLysGluAspCysLysAsp 894
Db      2557 TTGAAAGAAATATATATATCATCTCACTGAAGAAATGTGTTCAAAAAAGAGTGT---GAC 2613
Qy      895 LeuGluIleuValCysLysLysIleAsnLysAlaCysArgAsnLeuLysProIleuGluVal 914
Db      2614 TGTGAAGATTCATGTAATAAGATACAGAGAAATATGTTCAACATTAGAGCCACTAAAGTG 2673
Qy      915 LysPheThrIleGluThrValThrGlu-----SerThrThrThrThrThrThrThrThr 932
Db      2674 AGACACACACGAAATAGTAACTTAAACAATCACTAACACACACACACACACACCAACT 2733
Qy      933 ThrValAlaAspProLysAlaThrGluCysLysSerLeuGluIleuThrAspThrTyrVal 952
Db      2734 ACCATTAAAGCGCAAGGCAACAGACTGCCACTCTTTACAGACAAAGATACGTGGGTC 2793
Qy      953 ThrGlnThrSerThrIleThrSerThrIleThrSerThrIleThrSerIle 972

```

```

Db      2794 ACAAAAGCTGAGCCCATCTAGACATCCAGAACCAACATCTACAGTCCAGTCAAGATA 2853
Qy      973 ThrIleuThrSerThrArgArgCysLysProThrLysCysThrThrGlyGluGluAspAsp 992
Db      2854 ACCTTGACCTTCGACAAACCGGTGTAGCTTACGAAAGTGTACGACAGAGAGAAAGATGA 2913
Qy      993 AlaGlyAspValyluPheSerGluGlyLeuArgMetSerGlyTyrAsnValMetArgGly 1012
Db      2914 GCAGGACCGTGAAGCCGAGTGAAGGGTTGAGATGAGTGAAGTGAAGTGAAGTGAAGGGG 2973
Qy      1013 ValIleValAlaMetValIleSerPheMetIle 1023
Db      2974 GTGTATTAGCAATGACGATTTCAATCATGATTT 3006

RESULT 15
PMCSGINS
LOCUS      3563 bp      mRNA      linear      PLN 30-JUL-1999
DEFINITION
Pneumocystis carinii mRNA for major surface glycoprotein, complete
cds.
ACCESSION      D21827
VERSION      D21827.1 GI:425784
KEYWORDS      major surface glycoprotein.
SOURCE      Pneumocystis carinii
ORGANISM      Pneumocystis carinii
Bukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE      1 (bases 1 to 3563)
AUTHORS      Kikada, K., Wada, M. and Nakamura, Y.
TITLE      Multi-gene family of major surface glycoproteins of Pneumocystis
carinii: full-size cDNA cloning and expression
JOURNAL      DNA Res. 1 (2), 57-66 (1994)
MEDLINE      96051981
PUBMED      7584029
REFERENCE      2 (bases 1 to 3563)
AUTHORS      Kikada, K.
TITLE      Direct Submission
JOURNAL      Submitted (27-OCT-1993) Kazuhiro Kikada, Institute of Medical
Science University of Tokyo, Department of Tumor Biology; 4-6-1
Shirokanedai, Minato-ku, Tokyo 108, Japan
(E-mail: kkitada@ngc.ims.u-tokyo.ac.jp, Tel: 03-3443-8111 (ex. 308),
Fax: 03-3443-6319)
COMMENT      Submitted (27-Oct-1993) to DDBJ by:
Kazuhiro Kikada
Department of Tumor Biology
Institute of Medical Science University of Tokyo
4-6-1 Shirokanedai, Minato-ku
Tokyo 108
Japan
Phone: 03-3443-8111 x308
Email: kkitada@ngc.ims.u-tokyo.ac.jp
Fax: 03-3443-6319.

FEATURES
Source
Location/Qualifiers
1..3563
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/strain="IMS-PC84"
/db_xref="taxon:4754"
/dev_stage="trophozoite"
333..3563
/codon_start=1
/product="major surface glycoprotein"
/protein_id="BAA04851.1"
/db_xref="GI:425785"
/translation="MAPPYRQAVQGAADIDIEKHLLAFIAKRYKEENKCKOPLYEY
CEBIEKIKIDGSDVNVKNVGI,CEBGKQDCKLKGVEKVLKAFEGILQELAKIDKDN
CKEYKEKICILIEBTDHGAIKENCVNLKEKYKLRKREKVAEELFRALGSAKDKCK
EKIEKYPLISRESIELMSFCLNPSGTCGLKGLGVCPLEKEDLDGSLAEKCHER
LEKHFYGEACNKAACKDPEKCKKNNITTKAPESDSPKPSASLIRIGPDVYVK
AEKGGIIGKTVGLDIPKSGSTPLDILLVLSDEKPTDIPKCTTALAEKCDASKYLD
HNLEKELDKNDKCKKELLDVNVKFERCTNLKLYIKDLSITFBRKAEKSDLSWGL
PTLFTGECBALESECFYLENACKDNEIGACQNLRSACYKKQGDRLNLFPOLELG
NLGI,VRYSRDPGCKKYVNGCTKLDKYVLRCLYPELICYGNSDIFLOSKEISALL
DDQRPFLERKDELEKKEKDELSDSLNLNEKICITLKRRCSEYFVTERPRKVFLEKED

```



```

Db      2085  ACTTATATAGAGAGAAAGAAATTCATTAGTGTTCATGTGCTTACAGAGAAACA 2144
Qy      617   Cysegluuewvettllyvsaervallysaaragcysegluvalpheyvlyvsaenllyle 636
Db      2145  TGTATGATATATGATGATTCACACAAAGTCMAAGATTGTATAATTTAAAGAAACATCAAG 2204
Qy      637   Alasertrylleileglupheugluvaenanthraenlyvilethrthleugluarg 656
Db      2205  AATGAGAAATTTGATGAGAAATTTGAAAGCAAAATGAAATGAAACCACTGAAAGA 2264
Qy      657   AsnCyarosertrpnlsthrtyrcyvaenlrgpbeserproaenCyvprogly---Leu 675
Db      2265  CTTTGACAAACATGAGGCGGCATGTGTCATCACTTGAGAACTGCGAAGTCCAGATCACTG 2324
Qy      676   Thrlysegluven-----SerCyvthrlysele---lyvlyvshleargluPro 690
Db      2325  AAAAAGAGAAATGGCAATATATCATATACAGACACATCGATGAAATAATGCAGAAACC 2384
Qy      691   PhetyrlysaarglyvalaleugluasrpalaleuvalyglueugllylyleuThr 710
Db      2385  TTTTAAAGATTGAA---TTAGAGATGATGATGATCTATCTGTTGAAAGGCAATTAAA 2441
Qy      711   AsrlysaerlyvCygluProalaleuylsaargtyrcys---Thrvalaleglyasvval 729
Db      2442  GATGAAATTTGATGATGAAATTTGAAAGCAATTTAGAAAGATATGCTCGATGTTGAAAGATGA 2501
Qy      730   AsnabnlaserllyeserlyleuCyelyvalaenThrlysaarpsenSerGlylyvser 749
Db      2502  GCATTTCAAACTCTGTATGTGTAATGTGATGATATCAACAGAGGAAT------ 2549
Qy      750   AsrpalubapalaaarglyvleuCysegluylvseuvalyvgluvalygluGlucys 769
Db      2550  -----GTTTGCACAAATTTAGTTGATTAAGTAAAGAAATGT 2588
Qy      770   LyenlaleuprothrgluuLeuglylnProalalaasrleuylvlysaertrylyserThr 789
Db      2589  CCTACTTTAAAAACCGATCTAGAGAGAGCAAAAAAGAGTTGAAAGTCAAGAAAGAA 2648
Qy      790   TyrgluuLeuylvlyvsaarglalegluvalametaenlyvserSerleuvalleuser 809
Db      2649  TACATGCGCTCAACAGCGCGCAGAAATCTCAAAAGAAAGCTAAGTTATGTCTATCA 2708
Qy      810   Leuilelyvsaenluserasnvalseerlysaerlysaenlyvsaenlyvsaen 829
Db      2709  AAGCTGGAAGAGCGCAATGCCAATGCCAGATGCCAGATGCCAGT-----GATTCTACA 2759
Qy      830   AlavaI----- 831
Db      2760  CTAGTACCGGCACACAAACAGCAACAGCAGGTCAATCAGATCAGGGTCAACAACA 2819
Qy      832  -----SerhenglyleuGlinaerThr----- 838
Db      2820  GTGCCACACACACACACAAATGGAACGCCAGACACACACAGATGGAACATCAGGCACA 2879
Qy      839  -----Thrlyvshis-----Vallyvleleu 845
Db      2880  CCAAGTGAACACCAAGACACAGGTCAACAGAGATTAATGCAGAACTTGACCTGTT 2939
Qy      846  ArgaarglyvalyvasrvalserValThrgluleugluvalalyvalapheasrleuvala 865
Db      2940  AAAAGGGCATGTATGATGAGAGATACAGAGACAGAGTAAAGCATTTGATGCAACG 2999
Qy      866  AlaeluvalpheyglyaargtyrvalasrleuylsegluatrgCyvaenlyvleuGluserAsp 885
Db      3000  ACGGTAGCATTTGAGATGTATTTGAAATTTGAAAGGAGATGCAGCGCTTTCAACTAGAT 3059
Qy      886  CybaarglileuylvsaerCyvlysaerpleuGlulvalCyvlyvlyvshleasrlyvala 905
Db      3060  TCGGCTTTAAGAGAGATGTCAAGAACTTAACAGAGCTTCCGAAATAATGACGAGTTA 3119
Qy      906  CybaargenleuylvProleuGlulvalyvsProhlsrlygluThrlyglu----- 922

```

```

Db      3120  TCGGAGAGAAATAAACCATTAGAAATTAAGCTTCATCATACAGAGCGCAAAAGAAATC 3179
Qy      923  -----SerThrThrThrThrThr----- 929
Db      3180  TCAACCACTAGACAGACCACTAGACGATTAACCAAACTACTACCAACTACGACG 3229
Qy      930  ThrThrThrThrValalaasrPro-----LysalaThr---GluCyvlyvser 944
Db      3240  ACAACTACTACTAACCAAGCCGGAAGTGAAGAAAGTAACAGAAAGATGTACATG 3299
Qy      945  LeuGlThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 964
Db      3300  ATACAAACCAACAGATACATGATGGAACCGGATCGATTCATGATGATGATGATGATGAT 3359
Qy      965  ThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 984
Db      3360  ACGTGCAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3419
Qy      985  CyvThrThr-----GlygluGlubapalaglyasrval 996
Db      3420  TGTACCAACCGATTCAACAAAGAGACAGATTAAGAGAGAGAGAGAGAGAGATGTA 3479
Qy      997  LysProsergluGluleuargMetSerGlyTTrphenValmetargglyvalilevalala 1016
Db      3480  AAACCGAATGATGGAGAAATTAAGATTCTGATATGATTAATAATTAATGTTGTGGA 3539
Qy      1017  MetValIleSerPheMetIle 1023
Db      3540  GTGATGTATGAGGATGATG 3560

```

Search completed: October 29, 2005, 13:09:25
 Job time : 10825 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: October 29, 2005, 02:46:00 ; Search time 1232 seconds

(without alignments)
4915.500 Million cell updates/sec

Title: US-10-654-416-14
Perfect score: 5411
Sequence: 1 MARAAVRQAAGTQNSIDEH.....MSGMNVKGVIVAMVTSFMI 1023

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=framed.p2n.model -DEV=x1h
-O=/cgn21/USPTO.spool/US10654416/runat.26102005.101021.21723/app_query.fasta.1.1159
-DB=Geneseq.1dDec04 -OPMT=fastcap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIG=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=p2n -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US10654416.OCGN.1.1-857/runat.26102005.101021.21723 -NCPUB=6 -ICPU=3
-NO_MMAP -ARGOQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq.1dDec04.*
1: geneseqn19808.*
2: geneseqn19908.*
3: geneseqn20008.*
4: geneseqn20018.*
5: geneseqn20028.*
6: geneseqn20038.*
7: geneseqn20048.*
8: geneseqn20058.*
9: geneseqn20068.*
10: geneseqn20078.*
11: geneseqn20088.*
12: geneseqn20098.*
13: geneseqn20108.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5411	100.0	3072	AA294069	AA294069 Pneumocys
2	4398.5	81.3	3054	AA294068	AA294068 Pneumocys
3	4352	80.4	3084	AA294066	AA294066 Pneumocys
4	3827	70.7	3090	AA294065	AA294065 Pneumocys
5	3401.5	62.9	3081	AA294067	AA294067 Pneumocys

6	2762.5	51.1	3042	AA294063	AA294063 Pneumocys
7	2404.5	44.4	3006	AA294064	AA294064 Pneumocys
8	1430	26.4	3521	AAQ41226	AAQ41226 Clone GP3
9	1057.5	19.5	3521	AAQ41230	AAQ41230 Gene enco
10	1052	19.4	2190	AAQ41225	AAQ41225 Clone PC1
11	993	18.4	2190	AAQ41223	AAQ41223 Clone PC3
12	944	17.4	2126	AAQ41229	AAQ41229 Clone GP1
13	913	16.9	2814	AAQ40201	AAQ40201 Sequence
14	912	16.9	2110	AAQ41228	AAQ41228 Clone GP2
15	859	15.9	2058	AAQ41227	AAQ41227 Clone GP4
16	769.5	14.2	1454	AAQ41224	AAQ41224 Clone PC5
17	557	10.3	1448	AAQ40202	AAQ40202 Sequence
18	396	7.3	249	AAQ4070	AAQ4070 P. Carini
19	331.5	6.1	5794	ACA34675	ACA34675 Prokaryot
20	308.5	5.7	5334	AAQ414598	AAQ414598 H. pylori
21	274	5.1	5373	ADSA6883	ADSA6883 Bacterial
22	274	5.1	5641	ABZ32359	ABZ32359 Candida a
23	273.5	5.1	3432	ACA27810	ACA27810 Prokaryot
24	272	5.0	2943	AAA70229	AAA70229 Plasmodiu
25	270.5	5.0	4087	ADSA4846	ADSA4846 Bacterial
26	270	5.0	5886	ADSA97707	ADSA97707 Rabbit a1
27	270	5.0	6010	ADQ22038	ADQ22038 Human sof
28	270	5.0	6010	ADQ17241	ADQ17241 Human sof
29	270	5.0	6010	ADQ17288	ADQ17288 Human sof
30	270	5.0	12801	ADSA97709	ADSA97709 Rabbit a1
31	269.5	5.0	4779	AAFS8751	AAFS8751 Nucleolar
32	269.5	5.0	4780	ABL64409	ABL64409 Stomach c
33	269	5.0	10170	ADBS8964	ADBS8964 Toxicity-
34	269	5.0	10170	ADBS3712	ADBS3712 Primary r
35	268	5.0	6016	AAH57390	AAH57390 Human bke
36	267.5	4.9	7985	ABL08391	ABL08391 Drosophila
37	266	4.9	6965	ADD29780	ADD29780 Human tum
38	266	4.9	7636	ACFS8150	ACFS8150 Human gon
39	266	4.9	7694	ABX04175	ABX04175 Human mRN
40	266	4.9	7694	ABX10347	ABX10347 DNA encod
41	266	4.9	7695	ABL68595	ABL68595 Kidney ca
42	266	4.9	7695	ACFS8151	ACFS8151 Human gon
43	266	4.9	7743	ACFS8152	ACFS8152 Human gon
44	266	4.9	7761	ACFS8153	ACFS8153 Human gon
45	266	4.9	8063	ADL63564	ADL63564 Human ova

ALIGNMENTS

RESULT 1	AA294069	standard; DNA; 3072 BP.
ID	AA294069	
XX	AA294069;	
AC	15-SEP-2003	(revised)
DT	05-JUN-2000	(first entry)
XX		
DE	Pneumocystis carinii major surface glycoprotein gene HMG35.	
XX		
KW	Major surface glycoprotein; MSG; HMG35; human; pneumonia; diagnosis; ds.	
XX		
OS	Pneumocystis carinii; sp. f. hominis.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..3072
FT		/*tag= a
FT		/product= "HMG35"
FT		/note= "a nucleic acid comprising residues 2821-3072 of this sequence is specifically claimed in Claim 28"
XX		
XX	MO200009760-A1.	
PD	24-FEB-2000:	
XX		
PF	17-AUG-1999;	99WO-US018750.
XX		
PR	17-AUG-1998;	98US-0096805P.

QY 521 AepGlnIysArgAepPheProThrAspIysAsnCyseIyGluLeuGlyAArgIysCyseGln 540
 DB 1561 GATCAAAAGCAGATTTCCCGACAGATTAATAATTCAGAGAAATGGGGAGAAATGCCCA 1620
 QY 541 AepLeuGlyIysAepSerIyGluIleThrTrpProCyshIsthLeuGluGlnIys 560
 DB 1621 GATTAGAGAGATTCMAAAGAAATTCATGGCCATGTCATACATCGAGACCAATATGC 1680
 QY 561 AspArgLeuGlyIleThrTrpGluIleLeuIyGluValLeuLeuAepGluIleIysAepThr 580
 DB 1681 AATGCTTTGGGAGCTACAGAAATTTTAAAGCAGGTTTATGATGAACAAAGAAATCT 1740
 QY 581 LeuIysAepGlnIysCyseValIyEtyrLeuIyGluIysCyseAsnIyTrpSerArg 600
 DB 1741 TTGAAAGACCAAGAAAGTTGCTAAATTACTTAAAGAAAGGTATTAATGATCTAGA 1800
 QY 601 ArgGlyAspAepArgPheSerPheValCyseValPheGlnAsnAlaThrCyseGluLeuMet 620
 DB 1801 AGAGAGATGACCGTTCTCTTTGTATGTGCTTCCAAACGCTACGATGAGCTGATG 1860
 QY 621 ValIysAepValIysAepArgCyseGluValPheIyIysAsnIleIyValIleSerTyrIle 640
 DB 1861 GTAAAGACGTGAAGACAGGTGAGATTCAAAAAATATATTAAGCTTCATATATT 1920
 QY 641 IleGluPheLeuGluAsnAsnThrAsnIyIleThrTrpLeuGluAArgAsnCyseProSer 660
 DB 1921 ATTGAATTTCTTGAAATTAATACAAATTAATTAACACACTCGAAAGAAATTTGCCCTCT 1980
 QY 661 TrpIleThrTyrCyseAsnArgPheSerProAsnCyseProGlyLeuThrIyGluAsnSer 680
 DB 1981 TGGCATATCGATTCATATAGATTTTCACTTAATGTCAGAGTCTTACGAAAGAAATATGT 2040
 QY 681 CyethrIyIleIy 700
 DB 2041 TGTCAAAAATTCAGAAAGCATCGAGCGCTTCTATTAAGAAAGCGCTTGGAAGATGCT 2100
 QY 701 LeuIy 720
 DB 2101 CTCAAGTAGAGCTTCAGAGAAATTTGACTATTAATTTAAATGTAACCTCGCAATTGAAA 2160
 QY 721 ArgTyrCyethrValAlaGlyAsnValAsnAsnIleSerGlyLeuCyseIyValAla 740
 DB 2161 AGATATTTGTACAGTAGCGGAAACGTAAATATGTCATCAATCAATGCGTTATGCAAGCT 2220
 QY 741 AsnThrIyAspAsnSerGlyIy 760
 DB 2221 AACACCAAGATTAATCTCGAAAGATGATGAGATGCTTAAGAAAGAACTCTGTGAGAAA 2280
 QY 761 LeuValIy 780
 DB 2281 TTATGTGAAGAGGAGAAAGACAGTCAGAACCATTTACCAACAGAAATTTAGCAACCGGCA 2340
 QY 781 AlaAepLeuIy 800
 DB 2341 GCTGATCTAAAAAAGATTAATAGACATAGAGAACTTAAGAAACGTGCGAGAGAAACA 2400
 QY 801 MetAsnIy 820
 DB 2401 ATGAACAAGTCCAGCTTTGTTGTCTCACTTAAGAAAAAGAAAGAAAGTATCAAAA 2460
 QY 821 SerAsnSerIyAsnIyAspIyAsnAlaValSerAsnIyLeuGlnIyAsnThrThrIyIy 840
 DB 2461 AGTATATGCAAAAAAGATTAAGATGCGTTTCAACGGAATTCAGAGATTCACAAAA 2520
 QY 841 HisValIy 860
 DB 2521 CATGTGAAAAATACATACGAGGAGGAGTTAAGATGATCGATACAGAAATTTGAAGCTTAA 2580
 QY 861 AlaPheAepLeuAlaIleGluValPheGlyAArgTyrValIleAepLeuIyGluIyCyseAsn 880
 DB 2581 GCATTTGATTTGGCAGCAGAAATTTTGAAGATATGTATTTGAAGAAAGATGTAAAT 2640

QY 881 IyIyLeuGluSerAspCyseArgIleIyGluAspCyseIyAspLeuGluValIyCyseIy 900
 DB 2641 AAATTTGAATCAAGATTCAGATTAAGAGAGATTTGCAAGACTTAAAGAAAGTATGCAAA 2700
 QY 901 IyIyIleAsnIy 920
 DB 2701 AAGATTAATAGGCTTGTCCCAATCTGAAGCTTCGAGGTGAAGCGGACGAAACAGTG 2760
 QY 921 ThrGluSerThrThrThrThrThrThrThrThrThrThrValAlaAepProIyValIleThr 940
 DB 2761 ACAGAAAGTACACAGACAACTACACAAACAAACAAACCGTCCGATCCGAGGCAACG 2820
 QY 941 GluCyseIySerLeuGlnIy 960
 DB 2821 GAATGCAATCTTACAGACAAACAGACATGAGGTTCACAGACATGACACACAAAGC 2880
 QY 961 ThrSerThrIleThrSerThrIleThrSerIyIleThrIleThrIyIyIyIyIyIyIyIy 980
 DB 2881 ACGTCTACTATCATCTACCATCAATCAAAATTAATTAATTAATTAATTAATTAATTAAT 2940
 QY 981 IyIyProThrIy 1000
 DB 2941 AAACCAACCAAGTATGACAGAGAGAGAGATGATCAGAGACGTGAAACCAAGTAG 3000
 QY 1001 GlyLeuArgMetSerGlyTTPAsnValMetArgIyValIleValAlaMetValIleSer 1020
 DB 3001 GGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3060
 QY 1021 PheMetIle 1023
 DB 3061 TTCATGATT 3069
 DB
 RESULT 2
 AA294068
 ID AA294068 standard; DNA; 3054 BP.
 XX
 AC AA294068;
 XX
 DT 15-SEP-2003 (revised)
 DT 05-JUN-2000 (first entry)
 XX
 DE Pneumocystis carinii major surface glycoprotein gene HMSG33.
 XX
 KW Major surface glycoprotein; MSG; HMSG33; human; pneumonia; diagnosis; ds.
 XX
 OS Pneumocystis carinii; sp. f. hominis.
 XX
 Key Location/Qualifiers
 FT CDS 1..3054
 FT /*frag= a
 FT /product= "HMSG33"
 FT /note= "a nucleic acid comprising residues 2887-3132 of
 FT this sequence is specifically claimed in Claim 28"
 FT
 MO200009760-A1.
 PD 24-FEB-2000.
 XX
 XX 17-AUG-1999; 99WC-US018750.
 XX
 XX 17-AUG-1998; 98US-0096805P.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 XX Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;
 DR WPI; 2000-206025/18.
 DR P-PSDB; AAY79170.
 XX
 XX Detection of the presence of Pneumocystis carinii in specimens by
 PT identification of major surface glycoprotein (MSG) gene sequences using
 PT two or more oligonucleotide primers derived from human P. carinii MSG
 PT protein encoding sequence.

Qy	576	GIUHIIElybArPThrIleuLybAerGIuGuseCyvAlLyvTyIteuLybGIuLyCyv	595
Db	1714	GAACACAAAGTACTTGAAGACCAAGAAAGTGTGTAAATATACCTTAAAGAAAGTGT	1773
Qy	596	AenLyETpSerATgATgGIyAbpAbpArpPheSerPheValCyvAlPheGIuAbnAlA	615
Db	1774	AAATAATAGTCTTAGAGAGAGAGATGACCGCTTTCTTTTGTATGTGTCTTCCAAACGCT	1833
Qy	616	ThrCyvGIuLeuMeValLybArbValLybArbArGcyvGIuValPheLybLybAbnIle	635
Db	1834	ACGGTGAGCTGAAGTGTAAAGACGTGAAGACAGGTGTGAAGTATTCAAAATAATA	1893
Qy	636	LybAlaSerTyIleIleGIuPheLeuGIuAbnAbnThrAbnLybIleThrThLeuGIu	655
Db	1894	AAACCTTCATATATATTGAAATTTCTTGAAATATAATACAAATAATACAACTGGAA	1953
Qy	656	ArgAbnCyvPProSerThrPheIleThrTyCyvAbnArGpPheSerProAbnCyvProGIuLeu	675
Db	1954	AGAAATGTCCCTCTTGCGCATACGATATGATATGATTTCACTCAATATGTCACAGCTCT	2013
Qy	676	ThrLybGIuAbnSerCyvThrLybIleLybLybAbnGIuPProPheTyLybArGyA	695
Db	2014	ACGAAAGGAATAGTGTACAAAATCAAGAAAGCTGTGAGCCGTTCTTAAAGAAAG	2073
Qy	696	AlaLeuGIuAbpAlaLeuLybValGIuLeuGIuGIuLybLeuThrAbpLybSerLybCyv	715
Db	2074	GCTTTGGAAAGATGCTCTCAAGTAGAGCTTCAAGAAATTTGATGATTAATCTTAATGT	2133
Qy	716	GIuPProAlaLeuLybArGTyCyvThrValAlaGIuAbnValAbnAbnAlaSerIleSer	735
Db	2134	GAACCTGCATTTGAAAGATATTTGTACATACGCGGAAACGTAAATTAAGCCGTCAATCACT	2193
Qy	736	GIuLeuCyvLybAlaAbnThrLybArbAbnSerGIuLybSerArbGIuAbpAlaArgLyA	755
Db	2194	GGCTTATGCCAAAGCTTACACCAAGATATCTTGGAAAGATGATGAGATCTTAGAAG	2253
Qy	756	GIuLeuCyvGIuLybLeuValLybGIuValGIuGIuGIuGIuCyvLybAlaLeuProThrGIu	775
Db	2254	GAACCTGTGTGAAGAAATTAGTAAAGAAAGTGAAGAAAGCTGCAAGCTTACCAACAGAA	2313
Qy	776	LeuGIuGIuPProAlaAlaAbpLeuLybLybArbTyIleThrThrTyGIuGIuLeuLybLyA	795
Db	2314	TTAGGACAAACGGAGCTGATTTTAAAGAAAGTTTAAAGCATATGAGAACTTAGAA	2373
Qy	796	ArgAlaGIuGIuAlaMeTAbnLybSerSerLeuValLeuSerLeuIleLybLybAbnGIu	815
Db	2374	CGTCGAGAGGAAGCAATGACAAAGTCCAGTCTTGTCTGCTCATCTTAAGAAAAAGAA	2433
Qy	816	SerAbnValSerLybSerAbnSerLybAbnLybAbnLybAbnAlaValSerAbnGIuLeu	835
Db	2434	AGTATATGTATCAAAAAGTAAATAGCAAAAACAAGATTAAGATGCCGTTTCAACCGACTT	2493
Qy	836	GIuAbpThrThrLybAbnValLybIleLeuArGArGGIuValLybArbValSerValThr	855
Db	2494	CAAGATCACCAAAACATGTGAAAATCTACGAGAGAGAGATTAAAGATGTATCCGTAAACA	2553
Qy	856	GIuLeuGIuAlaLybAlaPheAbpLeuAlaAlaGIuValPheGIuArGTyValAbpLeu	875
Db	2554	GAATTAGAGCTTAAAGCAATTTGATTTGCCACAGAGATTTGGAAAGATATGTAGATTTGG	2613
Qy	876	LybGIuArGcyvAbnLybLeuGIuSerArpCyvArGIuIleLybGIuAbpCyvLybArbLeu	895
Db	2614	AAAGAAAGATGTAAATTAATTTGAATTCAGATTTGCAAAATTAAGAGAGATTGCAAAAGCTTA	2673
Qy	896	GIuGIuValaCyvLybLybIleAbnLybAlaCyvArGAbnLeuLybProLeuGIuValLyA	915
Db	2674	GAAGAAAGATGTAAAAAGATTATAAGCTTGTGCCAATCTGAAGCCTCTGGAGGTGAAG	2733
Qy	916	ProHIEGIuThrValThrGIuSerThrThrThrThrThrThrThrThrThrThrValAla	935
Db	2734	CCGACCGAAACAGTGCAGAAAGTATCAACGACATCTTACAAACAACAACACGCTTCC	2793

Oy		936	AspProlyValIaThrGluCysLysSerLeuGlnThrThraSPHrTPVallThrInThr	955
Dd		2794	GATCGAAGGAACGGAGTGCAAATCTTACAGACAACAACATGGGTTCACAGACA	2853
Oy		956	SerThrHisThrSerThrSerThrIleThrSerThrIleThrSerLysIleThrLeuThr	975
Dd		2854	TCCGACACACCAAGCATGTTACTACATTCATTCATTCATCATCAAAAATAAACATTGACA	2913
Oy		976	SerThrArgArgCysLysPProThrLysCysThrThriGlyGluIuaapaPaIlaGlyAsp	995
Dd		2914	TCAACGAGGGGATGCCAACCAACCAAGTGCACGACGGG-----GATGATGCGAGAAGAC	2967
Oy		996	VallyProSerGluGluLysArgMetSerGlyTrpAnaValMetArgGlyValIleVal	1015
Dd		2968	GTGAAGCCAAGTAGAAGCTTGAGGGTGAAGCGGGTGAATGATGAGGGGGGTGATAGTA	3027
Oy		1016	AlaMetValIleSerPheMetIle 1023	
Dd		3028	GCATGTGTTATTTCGTTTCATGATT 3051	
RESULT 3				
AAZ94066	ID	AAZ94066	standard; DNA; 3084 BP.	
AAZ94066;	AC			
15-SEP-2003	DT	(revised)		
05-JUN-2000	DT	(first entry)		
Pneumocystis carinii major surface glycoprotein gene HMSG14.	XX			
Major surface glycoprotein; MSG, HMSG14; human; pneumonia; diagnosis; ds.	KW			
Pneumocystis carinii; sp. f. hominis.	OS			
Location/Qualifiers	FH			
1..3084	FT			
/*tag= a	FT			
/product= "HMSG14"	FT			
/note= "a nucleic acid comprising residues 2839-3084 of	FT			
this sequence is specifically claimed in Claim 28"	FT			
WO200009760-A1.	FN			
24-FEB-2000.	PD			
17-AUG-1999; 99WO-US018750.	PP			
17-AUG-1998; 98US-0096805P.	PR			
(USSH) US DEPT HEALTH & HUMAN SERVICES.	PA			
Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;	PI			
WP1: 2000-206025/18.	DR			
P-PSDB; AAY79168.	DR			
Detection of the presence of Pneumocystis carinii in specimens by	PT			
identification of major surface glycoprotein (MSG) gene sequences using	PT			
two or more oligonucleotide primers derived from human P. carinii MSG	PT			
protein encoding sequence.	PT			
Claim 27; Page 72-76; 110pp; English.	PS			
The present sequence is that of the novel Pneumocystis carinii sp. f.	XX			
hominis gene, HMSG14, which encodes a major surface glycoprotein (MSG,	CC			
see AAY79168). The gene was isolated by PCR amplification of DNA taken	CC			
from an autopsy lung sample of an HIV-infected patient with P. carinii	CC			
pneumonia. It is 1 of 7 novel, claimed genes (see AAY79165-71). The MSGs	CC			
invention that encode human-P. carinii MSGs (see AAY79165-71). The MSGs	CC			
include a highly conserved C-terminal region of approximately 100 amino	CC			
acids; this region (also claimed) corresponds to residues 2839-3084 of	CC			
the present sequence. Direct detection or amplification of human-P.	CC			

CC carini1 MSG-encoding genes, especially by PCR using primers directed at
 CC the conserved region of the genes, provides a sensitive and specific
 CC technique for the detection of P. carinii, and the diagnosis of P.
 CC carinii pneumonia, especially in biological specimens (e.g. blood,
 CC sputum) from immunocompromised patients such as those with HIV infection.
 CC (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 3084 BP, 1240 A, 455 C, 676 G, 713 T, 0 U, 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	3084
Score:	4352.00	Matches:	826
Percent Similarity:	87.28%	Conservative:	80
Best Local Similarity:	79.58%	Mismatches:	106
Query Match:	80.43%	Indels:	26
DB:	3	Gaps:	6

US-10-654-416-14 (1-1023) x AAZ94066 (1-3084)

OY 1 MetAlaATgAlaValLyAArgGlnAlaAlaGlyThrGlnAenSerTLeAspGluGluHis 20
 DB 1 ATGGCGCGGCGGCTCAGCGCAGCAAAAGGTGCAAGAAATGACAAATGACATGAGGAGCAT 60
 OY 21 ValLeuAlaLeuLleLeuLysGluAspGlyLeuSerGluGluGlnCysAlaGlyLeu 40
 DB 61 GTTTAGCTTGATTTTAAAAAATGATTGAGAAATGACAAATGACAAATCTAAGTTG 120
 OY 41 LysLysTyrCysGlnGluLeuThrGlnAlaLysLysAenLleGluGlnValHisArgLys 60
 DB 121 GAGGAATTTGCAAAACATTAACAATGACAGATTAAATCCAGAAAAGTTACACAAA 180
 OY 61 LeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAlaAsn 80
 DB 181 TTAAGAATTTCTGTGATACCGGAAACGAATGATAAATGCAAGATCTAAACAA 240
 OY 81 IleGluLysLysCysThrThrLleLysGlyLysLysLysGluAlaLleLysLysLysLle 100
 DB 241 GTCAATCAAAATGCAATTAATTTTCAAGGAAACCTTCAACAGCTGCTAGAAAAAAT 300
 OY 101 GlnLleLleThrAspLysAspCysLysGluAsnGluGlnGlnCysLysPheLeuGluGly 120
 DB 301 TCAGAAATTAACAGATGAGGATTGCAAAAAGATGAACAACAATGCTATTTTTGGAGGA 360
 OY 121 ValCysSerLysGluLeuLysAspLysAspThrLysGluAsnLysCysTyrGlnLys 140
 DB 361 GCATGTCCAAACAGAACTTAAGATGACCTGCAATAATTAAGCAATGATGTTATCAAAA 420
 OY 141 LysATgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGly 160
 DB 421 GAACGAAACAATGGCAGAAAGATTTCTTTGAGGCGGCTTCGTGTGATCTCAATGA 480
 OY 161 SerValLleCysGluLysLysLysLysLysLysLysCysProValMetGlyAArgLysAsp 180
 DB 481 ACAAGACATGTGAAAAAAGTGAAGAAGTTGCCGAATTTGAAGAAGAAAGCAAT 540
 OY 181 GluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnLleLeuLleGluLys 200
 DB 541 GAATTAACGAGCTTGTCTTATCAAAAAACACATGCGTAAGCTTGTAAACAAAGCA 600
 OY 201 AspLysLysCysGlyThrLysLysThrAspValSerAlaLleLeuGlySerPheLys--- 219
 DB 601 AAAAGTAATGTGATCTTGAAGAAAGTTGAAGAACCTTAAGAACATCAATTTG 660
 OY 220 LysGluThrCysLeuGluLeuLeuGlnCysTyrPheTyrLleGlyAsnCysGlyAsp 239
 DB 661 CGAGAAAATGTCTATTAATCTTGAACAATGTTACTTTCAACAGAGGAACGTGAAGA 720
 OY 240 AspAsp-----LleLleLysCys 245
 DB 721 GACCAATCAAAAGTCAATTAACCTAATAATAAGACCTGCAAGAAATATGATACCAAGTGT 780
 OY 246 IleGluLeuGlyGlyLysCysGlnGluGlnAenLleAlaTyrMetProGlyProAsp 265

DB 781 GATCAATTAGCAAAAAGTGTGAAAAAATAATTGTTATATGATCATCCAGATCCGAT 840
 OY 266 PheAspProThrAspProGluAlaThrLleAlaGluAspLleGlyLeuGluPheTyr 285
 DB 841 TTGCATCAACTAAGCCAGAGCTTACACTGACAGAGACATAGGGCTGGAAGACTTAT 900
 OY 286 LysLysValGluGluAspGlyValPheLleGlyLysAsnHisLysAlaAspAlaThrAla 305
 DB 901 AAGAGGCGAGAAAGGATGGAATTTTGTGAAGACAAACGTGTAAGATGACAAACGCT 960
 OY 306 LeuLeuAlaLeuLleGlnAspSerSerLeuLysLysAspAspLysGluLysCys 325
 DB 961 TTGTGGCACTACTT-----CTTAAGAAAACCTTAAGAAAAGAAAGT 1005
 OY 326 GluGluAlaLeuGlnLysSerCysLysAsnProHisGluHisGluAlaLeuGluSerLeu 345
 DB 1006 ATAAAGCCCTTAATAAAAAAAGCTGCAAAACCTCATGAACATGAGGCTTGAGAAATCTA 1065
 OY 346 CysLysLysAsnGlyLeuSerAsnAspGlyThrLysLysCysGluGluLeuGlnAspAsp 365
 DB 1066 TGTAAAGAAATTAACCAAGTATGATGAAACGAAAAATGTGATGAATGAAAAAGAT 1125
 OY 366 IleAsnLysThrCysLysLlePheThrSerLysValThrAsnAsnArgLeuPheAspPro 385
 DB 1126 GTTAACAAAACCTTATCAAGCTTACATCAACAATTTCTTAATAAACCGCTTTACATTCA 1185
 OY 386 ThrLysGlyAsnAsnGlnLleValGlyTyrPglLysLeuProThrPheLeuSerAsnGlu 405
 DB 1186 CCTGATGA-----ATTGCGGAATGGGGAATAATTAACGACATTTCTTAGTGATA 1236
 OY 406 AspCysAlaLysLeuGluSerTyrCysPheTyrPheGluLysLysCysProAspGlyGlu 425
 DB 1237 GATTGTGCAAAACCTGAATCTTATTTGCTTATTAATAAGAACTTGCCAGATGTCAA 1296
 OY 426 AsnAlaCysLysAsnLleArgAlaThrCysTyrLysArgGlyLysAspAlaArgAlaAsn 445
 DB 1297 GAAGCTGTATGAATGTGAGGCGAGCGCTTTACAAAGAGGCGCTTGATGACGCGCAAC 1356
 OY 446 LysValLeuGlnGluAsnMetArgLysLysLysLysLysLysLysSerAspLysSerTyrPglLys 465
 DB 1357 AGTGTTGCAAAAATAATATGCGGTATTAATGATGATTAATCAAAATGCGCTTAAG 1416
 OY 466 LysPheGlnGlnGluLeuValLysValCysGluLysLysLysLysLysLysLysLys 485
 DB 1417 GAGTTTCAACAAAGATTATGAATAAGTATGTAAGAGCTA---AAAGAAATAAAGGAAGT 1473
 OY 486 PheSerAsnAspGluLeuPheLleLeuCysValGlnProAlaLysAlaAlaArgLeuLeu 505
 DB 1474 TTCCCAACGATGAATATTTTGTCTGTGTATACGCCACGAAAGCTGCACGATTACTT 1533
 OY 506 ThrHisAspLeuArgMetLysThrLlePheLeuArgGlnGlnLeuAspGlnLysArgAsp 525
 DB 1534 ACACACGATCATCAATAGAGGTTACCTTTTACGACAAACATTTGATCAAAAGAGAGAT 1593
 OY 526 PheProThrAspLysAsnCysLysGluLeuGlnLysArgLysCysGlnAspLeuGlyGluAsp 545
 DB 1594 TTTCCGACAGATTAAGACTGCAAGAACTAGGAAAAAATCCCAAGATTTAGGAAAGAT 1653
 OY 546 SerLysGluLleThrTyrProCysHisThrLeuGlnGlnGlnCysAsnArgLeuGlyThr 565
 DB 1654 TCAAAAGAAATTAACATGCGCATGTCTACACTGAGCGACGACATCAACGCTTGGGACT 1713
 OY 566 ThrGluLleLeuLysGlnValLeuLeuAspGluHisLysAspThrLeuLysAspGlnLys 585
 DB 1714 ACAGAAATTTAAAGCAGGTTTATTTGAGTGAACAAACAAAGATCTTGAAGACCAAGAA 1773
 OY 586 SerCysValLysTyrLysLysGluLysCysAsnLysTyrPserArgArgGlyAspAspArg 605
 DB 1774 AGTGTGTAATAATCTTAAGAAAGATGATATTAATGCTTAGAAGGAGATGACCGT 1833
 OY 606 PheSerPheValCysValPheGlnAsnAlaThrCysGluLeuMetValLysAspValLys 625
 DB 1834 TTCTTTTGTATGTGTTTTCAAAACGCTTACGTGTAGCTGATGTTAAAAAGACGTGAAA 1893

QY 626 AaPaRgCySgIuVa1PheLyLeAaSn1LeYsAlaSeRtYrIleIleGluPheLeuGlu 645
 Db 1894 GACAGGTGTAAGATTCAAAABAAAATTAAGACCTTCATATTAATTGAATTTCTTGA 1953
 QY 646 AaPaAnThrAaSn1LeThThLeuGluAaRgAaCyPSeRSeTTPHLeThTyCyS 665
 Db 1954 AATATATCAATATAAATMAACAACGCGAAGAAATGTCCTTGGCATATACGATATG 2013
 QY 666 AaPaRgPheSeRPeRoAnCyPProG1YLeuThrLySg1UaSnSeRtYrThLyIleLyS 685
 Db 2014 AATAGATTTTCACCTAATGTCAGGCTTACGAAAGAAATAGTGTACAAAATCAAG 2073
 QY 686 LyS1LeAaRgIuPProPheTyZLySArGlySAlaLeuGluAaPaLaLeuLyVaIleu 705
 Db 2074 AAGCATGTGAGCCGTTCTAATAAAGAAAGCCCTTGAAAGTGTCTCAAAAGTAGACT 2133
 QY 706 GlnG1YLyLeuThrAaPlySeRtYrCySgIuPProAlaLeuLyAaRtYrCySThVa1 725
 Db 2134 CAAGGAAATAGCATTAATCTAAATGTGAACCTGCATGAACAGATATTGTACAGTA 2193
 QY 726 AlaG1YLeuVa1AaPaAnAlaSeRtYrLeuCySlySAlaAaThrLySArPaSn 745
 Db 2194 GCGGAAACGTAATATGCTCATACAGTGGCTTATGCAAAAGCTAACACCAAGATPAC 2253
 QY 746 SeRtYrLySeSeRPaRgIuAaPaAlaRgLySgIuLeuCySgIuLySLeuVa1LyGluVa1 765
 Db 2254 TCTCGAAAGAGTGAAGAGTCTAGAGAGAACTCTGTGAAATCACTGAAAGAAAGTG 2313
 QY 766 GluG1uG1nCySlySAlaLeuPProThrG1uLeuG1Yg1nProAlaA1AaPLeuLyS 785
 Db 2314 GAAGAACAGTGCAAGCATTCACAAAGAAATTAAGACAAACCGGACGCTGATCTAAAAAA 2373
 QY 786 AaPtyrLySeThTyRgIuGluLeuLySArG1aG1uG1uAlaMeTaSnLySeSeR 805
 Db 2374 GATTATAGACATATGAGGAACTTAAGAAACGTGACAGGAAAGCAATGAACCTCAGT 2433
 QY 806 LeuVa1LeuSeRLeuIleYLySArG1uSeRAnVa1SeRtYrSeRSeRSeRSeR 825
 Db 2434 CTGTTTGTGACCTCATTAAGAAABAAACGAAGTAAATGTAAGAAAGTAAATGCAAAAC 2493
 QY 826 LySArPlySArAlaVa1SeRAnG1YLeuG1aAaPThrLySAlaLyS1Leu 845
 Db 2494 AAGGATTAAGATGCCGTTTCAACGAGCTTCAAGATTCACAAACATGTGAAATACTA 2553
 QY 846 ArGArG1YVa1LySArPaVa1SeRValThrG1uLeuG1uAlaLySAlaPheAaPLeuAla 865
 Db 2554 CGGAGAGAGATTAAAGATGATCCGTAAACAAATTAAGAACTTAAGCATTTGATTGGCA 2613
 QY 866 AlaG1uVa1PheG1YArGtYrVa1AaPLeuLySgIuAaRgCySArSn1LySLeuG1SeRPa 885
 Db 2614 GCAGAAATATTTGAGAAATATGTAATTTGAAGAAAGATTAATTAATTTGGAATTCGAT 2673
 QY 886 CySArG1YLeuSgIuAaPcyLySArPLeuGluGluVa1CySlyS1LeuLySAla 905
 Db 2674 TGCGAATTTAAAGAGATTGCAAAAGCTTAAGAAAGATTAAGAAAGATTAATTAAGGCT 2733
 QY 906 CySArGArSnLeuLySProLeuGluVa1LySProH1eG1uThrVa1ThrG1uSeRThThr 925
 Db 2734 TGTCGCAATCGAACCTCTGAGAGTAAAGCCGACCAACAAACAGACAGAAAGTACACAG 2793
 QY 926 ThThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 945
 Db 2794 ACAACTCAACACAAACAAACAAACCTTGCCGATCCGAAGGCAAGCAATGCAATCTCTTA 2853
 QY 946 GlnThrThrAaPThrTPVa1ThG1nThSeRThrA1eThSeRThrSeRThrIleThr 965
 Db 2854 CAGACAAACAGACATGGGTACACAGACATCGACACACAAACAGCATCTACTATACACA 2913
 QY 966 SeRThrIleThrSeRtYrIleThLeuThSeRThrAaRgCySlySProThLyCyS 985
 Db 2914 TCTACCATCATCAATAAATAACATTAACATTAACAGAGCGATCAAAACCAACCAAGTGT 2973

QY 986 ThThrThrG1YLeuG1uAaPaPaAlaG1YAsPValLySProSeRtYrG1YLeuAaRgMeSeR 1005
 Db 2974 ACAGCAGG-----GATGATGCGAAGACGTGAAGCCAAAGTGAAGGCTTGAGGCGTACG 3027
 QY 1006 G1YTPAaSnVa1MeTaRgIyVa1IleVa1AlaMeTaVa1IleSeRPhMeIle 1023
 Db 3028 GGGTGAAATGATGATGAGGGGGGTGATGATGATGATGATGATGATGATGATGATGATG 3081
 RESULT 4
 AA294065
 ID AA294065 standard; DNA, 3090 BP.
 XX
 AC AA294065;
 XX
 DT 15-SEP-2003 (revised)
 DT 05-JUN-2000 (first entry)
 XX
 DE Pneumocystis carinii major surface glycoprotein gene HMSG11.
 XX
 KM Major surface glycoprotein; MSG; HMSG11; human; pneumonia; diagnosis; ds.
 XX
 OS Pneumocystis carinii; sp. f. hominis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..3090
 FT /*cag= a
 FT /product= "HMSG11"
 FT /note= "a nucleic acid comprising residues 2845-3090 of
 this sequence is specifically claimed in Claim 28"
 FT
 XX
 PN W0200009760-A1.
 XX
 PD 24-FEB-2000.
 XX
 PF 17-AUG-1999; 99MO-US018750.
 XX
 PR 17-AUG-1998; 98US-0096805P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Kovacs JA, Huang S, Maaur H, Fischer SH, Gill VJ, Wei Q;
 XX
 XX WPI: 2000-206025/18.
 DR P-PSDB; AAY79167.
 XX
 PT Detection of the presence of Pneumocystis carinii in specimens by
 PT identification of major surface glycoprotein (MSG) gene sequences using
 PT two or more oligonucleotide primers derived from human P. carinii MSG
 PT protein encoding sequence.
 XX
 PS Claim 27; Page 64-68; 110pp; English.
 XX
 CC The present sequence is that of the novel Pneumocystis carinii sp. f.
 CC hominis gene, HMSG11, which encodes a major surface glycoprotein (MSG,
 CC see AAY79167). The gene was isolated by PCR amplification of DNA taken
 CC from an autopsy lung sample of an HIV-infected patient with P. carinii
 CC pneumonia. It is 1 of 7 novel, claimed genes (see AA27963-69) of the
 CC invention that encode human-P. carinii MSGs (see AA79165-71). The MSGs
 CC include a highly conserved C-terminal region of approximately 100 amino
 CC acids; this region (also claimed) corresponds to residues 2845-3090 of
 CC the present sequence. Direct detection or amplification of human-P.
 CC carinii MSG-encoding genes, especially by PCR using primers directed at
 CC the conserved region of the genes, provides a sensitive and specific
 CC technique for the detection of P. carinii, and the diagnosis of P.
 CC carinii pneumonia, especially in biological specimens (e.g. blood,
 CC sputum) from immunocompromised patients such as those with HIV infection.
 CC (Updated on 15-SEP-2003 to standardise OS field)
 XX
 SQ Sequence 3090 BP; 1255 A; 447 C; 662 G; 726 T; 0 U; 0 Other;
 Alignment Scores: 3,18e-281 Length: 3090
 Pred. No.: 3827.00 Matches: 728
 Score:

Percent Similarity: 81.54% Conservative: 120
 Best Local Similarity: 70.00% Mismatches: 164
 Query Match: 70.73% Indels: 28
 Gaps: 8
 DB: 3
 US-10-654-416-14 (1-1023) x AA294065 (1-3090)

1 MetalaatglaValaLysaArglnAlaalegLyThrGlnaasrTlaaPgluGlnHs 20
 1 ATGGCGCGGCGGTCAACGCGCGGCAAAAGGTGACAGAAATGACATTTGATGAGAGCAT 60
 21 ValleaualaleuLleuLysaPgluAspGlyLeuSerGlnGlnGlnCysAluLysLysLeu 40
 61 GTTTAGCTTGATTTTAAAAAAAATGATTAAGAAATACAAAATGCAAACTAACTAAGTTG 120
 41 LysLysTyrcysGlnGlnLeuThrGlnAlaLysLeuAenllegLuglnValHlsargLys 60
 121 GAAGAATTTGCAAAACATTAAACAATGACAGATTAAATCCAGAAAAGTTGACGAAAATA 180
 61 LeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAlaAsn 80
 181 TTAAGAAATTTCTGTGATACCGGAAACGAATGAAAATGTCAGATCTAAACAAACAAA 240
 81 llegLugLysCysThrThrLleLysGlyLysLeuLysGlnAlaLleLysLysLyslle 100
 241 GTCAATCAAAAATGCAATTAAATTTCAAGGAAAATTTCAACAGCTGCTGGAATAAAATTT 300
 101 GlnllelIethrAspLysAspCysLysGlnaasnglnGlnCysLeuPheLuglnGly 120
 301 TCAGAAATTAACAGATGAGGATTGCAAAAAGAAATGAACAATGCTATTTTGGAGGGA 360
 121 ValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyGlnLys 140
 361 GCATGTCCAACAGAACTTAAAGATGACTGCATTAATTAAGAAATACCTGTTATCAAAA 420
 141 LysArgAspLysValAlaGlnGlnValleuLeuArgAlaLeuArgSerAspLeuAsnGly 160
 421 GAACGGAACATGTGCGAGAAAGATCTTTGAGGCGCTTCGCGTATCTCAATGA 480
 161 SerVallellecysGlnLysLysLeuLysGlnLysCysProvalMetClyAsnGlnSerAsp 180
 481 ACAAGACATGTGAATAAAGCTGAAGAAGTTGCCGAATTTGAAGAAGAAAGCAT 540
 181 GlnLeuThrAsnLeuCysLeuAsnGlnLysGlnLysCysLysAsnLleuLleGlnLys 200
 541 GAATTAACGAGCTTGTCTTATCAAAAACACATGCTGTAAGCTTGAACAAAAGA 600
 201 AspLysLysCysGlyThrLeuLysThrAspValSerAlaLleuGlnLysSerPheLys--- 219
 601 AAAAGTAATGTGATCTTGAATAAAGAAAGTTGAAGAAGCATTAAGAAAGAAATTTG 660
 220 LysGlnThrCysLeuGlnLysLeuLysGlnLysCysTyrcPheTyrllegLysAsnGlyAsp 239
 661 CGAGAAAATGTCTATTACTTGAAGCATGTTCCTTCAAGAGGGAACGTGAAAGA 720
 240 AspAsp-----lIleLysCys 245
 721 GACCAATCAAAAGTGCATTAACCTATATAAAGATGCAAGAAATATGACCAAGATGT 780
 246 lIlegLugLysGlyLysCysGlnGlnGlnAlaLysMetProGlyProAsp 265
 781 GATGAATTAAGCAAAAGGTGGAAGAAGAAATATGTTATATATCATCCAGAAATCCGAT 840
 266 PheAspProThrAspProGlnAlaThrIleAlaGlnAspLlegLysGlnLysPheTyrc 285
 841 TTTCATTCACAACTAACCGACGCTTACATGACAGAGACATAGGCTGGAAGAAGCTTAT 900
 286 LysLysValGlnGlnAspGlyValPheIlegLysAsnHlsLeuArgAspAlaThrAla 305
 901 AAGAGGCGAGAAAGATGAAATTTTGTGGAGAACAAACATGTAAAGATCAACACT 960
 306 LeuLeuAlaLeuLlellegLysAspSerLeuLysLysAspAspLysGlnLysCys 325

961 TTGTGGCACTACTT-----CTTAAGAAAACCTTAAGAAAAGAAATGT 1005
 326 GlnGlnAlaLeuGlnLysSerCysLysAsnProHlsGlnHlsGlnAlaLeuGlnSerLeu 345
 1006 ATAAAGCCCTTAATAAAAACTGCGAAAACCCCATGAAACATGAGGCTTGAATAATCTA 1065
 346 CysLysLysAsnGlyLeuSerAsnAspGlyThrLysLysCysGlnGlnLysGlnAspAsp 365
 1066 TGTAAAGAAATTAACCAAGATGATGAAACGAAAATGTGATGAACTAAGAAAAGAT 1125
 366 lIleAsnLysThrCysLysIlePheThrSerLysValThrAsnAlaArgLeuPheAspPro 385
 1126 GTTAACAAAACCTTACAGCTTACATCAACATTTCTTAATAAACCGCTTTACATTCA 1185
 386 ThrLysGlyAsnAsnGlnLleValGlyTyrcGlnGlyLeuProThrPheLeuSerAsnGln 405
 1186 CCTGATGA-----ATTGCGGAATGGGGAATAATTAACGACATTTCTTAAGTATGA 1236
 406 AspCysAlaLysLeuGlnLysSerTyrcPheThrPheGlnLysLysCysProAspGlyGln 425
 1237 GATTGTGCAAAACTAAGATCTTATGCTTTATTAAGAAACCTTGCCAGATGTCAA 1296
 426 AsnAlaCysLysAsnAlaArgAlaThrCysTyrcLysArgGlyLeuAspAlaArgAlaAsn 445
 1297 GAACCTGTATGAATGTGAGGCGACCGCTTATAAGAGAGGCTTGATGACGGGCAAAAC 1356
 446 LysValleuGlnGlnAsnMetArgGlyMetLeuHlsGlySerAsnLysSerTyrcGln 465
 1357 AGTGTTGTGCAAAAATATATGCGAGGTTATGTCATGCTCAAAATTAAGATTTGCTTAA 1416
 466 LysPheGlnGlnGlnLeuValLysValCysGlnLysLeuLysLysGlnAsnLysGlySer 485
 1417 AAATTTCAACAAAGATTAGCAAAAAGTATGTGAAGAACTG--AAAGAAATTAAGGAAAT 1473
 486 PheSerAsnAspGlnLeuPheIleLeuCysValGlnProAlaLysAlaLysArgLeuLeu 505
 1474 TTCTGCAACATGAATTTGTTCTGTGTATACACACAGAAAGGACGACGATTACTT 1533
 506 ThrHlsAspLeuArgMetLysThrIlePheLeuArgGlnGlnLeuAspGlnLysArgAsp 525
 1534 ACACATCACCATCAAAATGAGATTTATCTTTTGAACAACAACATGATCAAAAAGAGAT 1593
 526 PheProThrAspLysAsnCysLysGlnLysGlnLysArgLysCysGlnAsnLeuGlnLys 545
 1594 TTTCGACAGATTAAGACTGCAAGAAATTAAGGAAATATCCAAAGATTTGAGAAAGAT 1653
 546 SerLysGlnLleThrTyrcProCysHlsThrLeuGlnGlnGlnCysAsnArgLeuGlyThr 565
 1654 TCAAAAGAAATTAACATGCGCATGTCTATCACTAACAACGCAATGCAATCGCTTAGGAT 1713
 566 ThrGlnLleLeuLysGlnValleuLeuAspGlnHlsLysAspThrLeuLysAspGlnGln 585
 1714 ACAGAAATTTTAACAAGATTTTATGATGAACAACAAGATTAATTTGAAGAAGTCATGA 1773
 586 SerCysValLysTyrcLysLysGlnLysCysAsnLysTyrcPheArgGlnGlyAspAspArg 605
 1774 AACTGTGCAAAATTTTAAGAAATATGCAATTAAGTGTAAAGGGGTGAAGATGT 1833
 606 PheSerPheValCysValPheGlnAsnAlaThrCysGlnLysMetValLysAspValLys 625
 1834 TTTCTTTGTGATGTGTTTCCAAAACCTTCACTAGTGAAGCTGATGATAAGAAAGCTGCA 1893
 626 AspArgCysGlnValPheLysLysAsnLleLysAlaSerTyrcIlellegLysPheLugln 645
 1894 GATAGGTCAAAATATTTGAGAAATATGCAAGCATCAGATATTAATGATTCCTTA 1953
 646 AsnAsnThrAsnLysIleThrThrLeuGlnLysArgAsnCysProSerThrPheIleThrTyrc 665
 1954 AAAATATCAATTAAGCAAGATCAGCAACAAATATTTGTCCCTGATGCAATCCAAATG 2013
 666 AsnArgPheSerProAsnCysProGlyLeuThrLys---GlnAsnSerCysThrLysIle 684
 2014 GATGATTTTAAACCAATTTGCTGATCTTAAGAAAGAAAGAACTTTGTGCAAAATCTT 2073

```

QY 685 LyeIyehIaRgGluProPheTyRlyeAryyAlaLeuGluAaPaIaLeuLyVaIGlu 704
Db 2074 AAAAAAAAAATTCGCAACATTCCTACAAAAGAAAGTTTAAAGATGCTCTTAAAGTGAAG 2133
QY 705 LeuGInGlyLyLeuThrAAsPlySeRlyeCySgluProAlaLeuLyAaRyTyRyCyThr 724
Db 2134 CTTGAGAGAAATTTAAAGTAAATATATGAATGTGAACCTGATGATGAAAGATATTGTGACA 2193
QY 725 ValAlaGlyAenValAaenAaenAlaSerIleSeRgLyLeuCyLyVaIaAenThLyAaP 744
Db 2194 GTATTTGAAAGAGCTAAATTAATGCTCAATCAGCGATTATGTAAAGATTAATACCGAAAGT 2253
QY 745 AaenSeRgLySeSeRseP---GluAaPaIaRgLySeGluLeuCySgluLyLeuValLyS 763
Db 2254 AAAAACTAAAAAGCCGATTAATAAAAATCTTGAAAGAAAGCTTTGTCTTAAATTTGTGAA 2313
QY 764 GluVaIGluGluGInCyLyVaIaLeuProThRgIuLeuGlyGInProAlaIaAaPLeu 783
Db 2314 GAGGTGGAACGCAATGCAAGATTAATCAACAGAAATTAACAGAGCTGGAAGAAAGTCTA 2373
QY 784 LyLeLyAaPlyRlyeThRyTyRyGluGluLeuLyLyAaRgAlaGluGluAlaMeTaAaPlyS 803
Db 2374 AAAAAAGATGTTAAGACATATGAGAACTTAAGAAAGGCGCAAAAAAGCAATGACAAG 2433
QY 804 SeRSeRLeuValLeuSeRLeuIleLyLyAaenGluSeRSeRValSeRlyeSeRSeRSeR 823
Db 2434 TCCAGCTCTGTTTATACCTGTTTAAAGAAACGAAAGTAAATACATGCAAAAAATTAATAGC 2493
QY 824 LyAaenLyAaPlyAaenAlaValSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeR 843
Db 2494 AAAAAAGATGTTAAGATGCTGTTTCAACGCAATTAACAGTCAACAAAAATATGTGAAA 2553
QY 844 IleLeuAaRgATgLyVaLyAaPValSeRValThRgIuLeuGluAlaLyVaIaPheAaP 863
Db 2554 ATACTAGAAAGAGAGATTAAAGAGCACTTGTTAAGAAATGTGAAGCCAGCAAGCTTTGAT 2613
QY 864 LeuAlaIaGluValPheGlyAaRgTyRyValAaPLeuLyGluAaPlyAaenLyLeuGlu 883
Db 2614 TTGGCAGCAGAAAGCTTTGAAAGTATGTAGACTTTGAAAGAAATGTGAGAAATTTGACT 2673
QY 884 SeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeRSeR 903
Db 2674 TCGGATTCGGGAGTTAAAGACATTTGCGATGTTTAAAGAAAGTGTGGAAGATTTGAG 2733
QY 904 LyVaIaCyAaenAaenLeuLyPProLeuGluValLyPProIleGlyThRValThGluSeR 923
Db 2734 AAGCAGATGTCAGATCTGAAGCTCTGAGAGTGAAGTCGCAATGAATAGTCACAGAAAGC 2793
QY 924 ThRThRThRThRThRThRThRThRThRThRThRThRThRThRThRThRThRThRThR 943
Db 2794 ACAACGACGACACACACACACACACACACACACACACACACACACACACACACACACAC 2853
QY 944 SeRLeuGInThRThRThRThRThRThRThRThRThRThRThRThRThRThRThRThR 963
Db 2854 TCCTTACGAGAACAGATACATGAGTTTACACAGACATGACACACACACACACACACCTAAC 2913
QY 964 IleThRSeRThRThRThRThRThRThRThRThRThRThRThRThRThRThRThRThR 983
Db 2914 ATCAACATCTACCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2973
QY 984 LySeCyAaenThRThRThRThRThRThRThRThRThRThRThRThRThRThRThRThR 1003
Db 2974 AAGGTTCAGACAGG-----GATGAACGACAGAGACGTAAGGAGGAGGAGGAGGAGGAG 3027
QY 1004 MeSeRgLyTTPaenValMeTaRgLyValIleValaIaMeTaValIleSeRPhemeIle 1023
Db 3028 ATGATGGGTGAGAGCGATGATGAGGCGGTGATTAAGCAATGTTATTCCTCATGATT 3087

```

```

AC AA294067;
XX
DT 15-SEP-2003 (revised)
DT 05-JUN-2000 (first entry)
DE Pneumocystis carinii major surface glycoprotein gene HMSG32.
KM Major surface glycoprotein; MSG; HMSG32; human; pneumonia; diagnosis; ds.
XX
OS Pneumocystis carinii, sp. f. hominis.
XX
FH Key
FH CDS
FT 1. 3030
FT /*tag= a
FT /product= "HMSG32"
FT /note= "a nucleic acid comprising residues 2836-3081 of
FT this sequence is specifically claimed in Claim 28"
XX
XX W0200009760-A1.
XX
XX 24-FEB-2000.
XX
XX 17-AUG-1999; 99MO-US018750.
XX
XX 17-AUG-1998; 98US-0096805P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Kovacs JA, Huang S, Masur H, Fischer SH, Gill VI, Mei Q,
XX WPI: 2000-206025/18.
XX DR P-PSDB; AAY79169.
XX
XX
XX Detection of the presence of Pneumocystis carinii in specimens by
XX PT identification of major surface glycoprotein (MSG) gene sequences using
XX PT two or more oligonucleotide primers derived from human P. carinii MSG
XX PT protein encoding sequence.
XX
XX Claim 27; page 80-84; 110pp; English.
XX
XX The present sequence is that of the novel Pneumocystis carinii sp. f.
XX CC hominis gene, HMSG32, which encodes a major surface glycoprotein (MSG,
XX CC see AAY79169). The gene was isolated by PCR amplification of DNA taken
XX CC from an autopsy lung sample of an HIV-infected patient with P. carinii
XX CC pneumonia. It is 1 of 7 novel, claimed genes (see AA279063-69) of the
XX CC invention that encode human P. carinii MSGs (see AAY9165-71). The MSGs
XX CC include a highly conserved C-terminal region of approximately 100 amino
XX CC acids; this region (also claimed) corresponds to residues 2836-3081 of
XX CC the present sequence. Direct detection or amplification of human P.
XX CC carinii MSG-encoding genes, especially by PCR using primers directed at
XX CC the conserved region of the genes, provides a sensitive and specific
XX CC technique for the detection of P. carinii, and the diagnosis of P.
XX CC carinii pneumonia, especially in biological specimens (e.g. blood,
XX CC sputum) from immunocompromised patients such as those with HIV infection.
XX CC (Updated on 15-SEP-2003 to standardise OS field)
XX
XX
SQ Sequence 3081 BP; 1231 A; 431 C; 663 G; 756 T; 0 U; 0 Other;

```

```

Alignment Scores:
Pred. No.: 7.57e-249 Length: 3081
Score: 3401.50 Matches: 655
Percent Similarity: 76.17% Conservative: 141
Best Local Similarity: 62.68% Mismatches: 208
Query Match: 62.86% Indels: 41
DB: 3 Gaps: 12

```

```

US-10-654-416-14 (1-1023) x AA294067 (1-3081)
QY 1 MeTaIaRgAlaValLyAaRgGluAlaIaGlyThRgIuAaenSeR-----IleAaPglu 18
Db 1 ATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
QY 19 GluHISValLeuAlaLeuIleLeuLyGluAaPgluAaenSeRgInGluGluCyLyLyS 38

```

```

Db      61  GAAAGAAATCTTTGGCTTTAATTGCAGAAAAAATTAATGATTAATGAAATGCAAAAA 120
Qy      39  LybLeuLybLybCyseGlnGluLeuThrgluAlaLeuAenIleGluGlnValHis 58
Db      121 GAATTAGAAAAAATATGTATAGACATTTAAGCATGCAAGATTTAAACGAAAAAGTTAC 180
Qy      59  ArgLybLeuLybGlyPheCyseGlnAapGlyLysAlaAapThrLysCyseLybGluLeuLys 78
Db      181  AAAAAAATTAAGAGCTTTTGTGAAAAATAAAAAGCAGATTCAAAATGCAAAAGAACTGAAA 240
Qy      79  AlaAenIleGluLybLysCyseThrThrIleLybGlyLysLeuLybGluAlaIleLysLys 98
Db      241  GAAAAAATCACTCAAAAAATGTAATCTCAATCAAAAGAAAACTTACAAAGCAATCAAAAA 300
Qy      99  LysIleGlnIleIleThrAapLybAapCyseLybGlnAenGlnGlnCyseLeuPheLeu 118
Db      301  AAAAAATTCAGATTTAAGCATGATGCAAAAGAAATGAAACAAATGCTATTTTGG 360
Qy      119  GluGlyValCyseSerLybGluLeuLybAapCyseAapThrLysAapThrLysCyseLys 138
Db      361  GAGGAGCATGTCCAGCCGAACCTTAAGATGATTCAAATCTTTGAGAAATTAAGTCTAT 420
Qy      139  GlnLybLysAapLybAapLybValAlaGluGlnValLeuLeuArgAlaLeuArgSerAapLeu 158
Db      421  CAAAAAGAACGTGATTAAGTGGCGAAGAAAGCTCTTTTAAGAGCATGTCGAGCTTA 480
Qy      159  AapGlySerValIleCyseGlnLybLybLeuLybGlnIleCyseProValMetGlyArgGlu 178
Db      481  ATCAATGAACACTACATGTGAAAGAAAGCTCAAAAGAGTTTGATGAGTTCAGTCAAGAA 540
Qy      179  SerAapGluLeuThrAapLeuCyseLeuAenGlnLybGluThrCyseLysAenIleLeuIle 198
Db      541  AGTGTAGATTAACCAAGCTTTGTCTTATCAAAAAAGACCTGCAAAAACATTTGTATTA 600
Qy      199  GlnLybAapLybLysCyseGlyThrLeuLysThrAapValSerAlaIleLeuGlySerPhe 218
Db      601  GAAAAACAAAAAATGTAATGCTCTTAAACAGATGTTAAACGACGACCTTGAG----- 654
Qy      219  LysLybGlu-----ThrCyseLeuGluLeuGlnGlnCyseLysPheThrIle 234
Db      655  AAGAAAGATGAGTTACGAGAAATGTTTACACATGCTTGAACATGCTATTTATAGA 714
Qy      235  GLybAenCyseGlyAapAapAapIleIleLysCyseIleGlu----- 247
Db      715  GGGAAATTTGT-----GAAGATATATCAAAATGTAATTAATCATCCGAAGACTGTATGA 768
Qy      248  -----LeuGlyLysCyseGlnGlnGlnAenIleAThrMet 260
Db      769  TATTTGCCAGTGTGATATCAATTGGCAGATGAAATGAAAGAAATTAAGTTATTTATCA 828
Qy      261  ProProGlyProAapPheAapProThrArgProGluAlaThrIleAlaGluAapIleGly 280
Db      829  CATCGGGAGTCGATTTCAATCACTCAAGTCAAAAGCCATCTGTACGAGAAACATAGA 888
Qy      281  LeuGlnGluPheThrLysLybValAlaGluLysAapGlyValPheIleGlyLysAenHisLeu 300
Db      889  CTGGAAAGAGCTTTATTAATAAGCCCGCAGAAAGAGGTTCATATTTGAAAAGGCTCTCTGA 948
Qy      301  ArgAapAlaThrAlaLeuLeuAlaLeuLeuIleGlnAapSerSerLeuLybLysAap 320
Db      949  AAGATGCACTGCTACTGCGCTTTGATTCAAAATCTAAGATCTTAAGTGAAGTGAAGTG 1008
Qy      321  AapLybGlnLybCyseGlnGlnAlaLeuGlnLysSerCyseLysAapProHisIleGlnIle 340
Db      1009  GGTAAAGAA-----TGCGAAAAAATTTCTTAAGATTAAGTAAAGATTAAGAAAGTACATA 1065
Qy      341  AlaLeuGlnSerLeuCyseLysAenGlnLybSerAapAapGlyThrLysLybCyseGlu 360
Db      1066  ATTTTGGAGATTTTGTATTAATCAAAATGATGCTGTCAAAATGAATTTGAAAAGTGAATA 1125
Qy      361  GlnLeuGlnAapAapIleAapThrCyseLysIlePheThrSerLysValThrAenAen 380

```

```

Db      1126 GAGTTAAGAAAGAGTTAGCAAAACAGTACTAAATCTCTTTTGAAAAAATTAAGATTA 1185
Qy      381  ArgLeuPheAapProThrLysGlyAapAapGlnIleValGlyTrpGluLybProThr 400
Db      1186  CACCTC-----TCGGATCCGAGAAAGTCAATTCATCCATGGTATTAAGTACGACA 1233
Qy      401  PheLeuSerAenGlnAapCyseAlaLybLeuGlnLysSerThrCysePheThrPheGluLys 420
Db      1234  TTTTGTATGTCAAATGACTGCAAGAGTTAGATCAAGACTGTTTTTATTTAAAAAGTCAA 1293
Qy      421  CysePro--AapGlyGlnAenAlaCyseLysAenIleArgAlaThrCyseLysArgGly 439
Db      1294  GCACCTCTTGACAAAGAA-----TGTAATATCTGAAGCGACAGATGTTATTAAGAGCG 1347
Qy      440  LeuAapAlaArgAlaAenLybValLeuGlnGlnAenMetArgGlyMetLeuHisGlySer 459
Db      1348  CTTGAACACACAGTATTAAGACATTCAGAAAAAAGATGTACGAGACTGTTCTATAGTTCA 1407
Qy      460  AapLysSerThrLeuGlnLybPheGlnGlnLeuValLysValCyseGluLysLys 479
Db      1408  GGCAAAAGATGTTTAAGAACTTACTAAGAAAAATTAATGAAAGAAATGTTCCGAACCTTAA 1467
Qy      480  LybGluAapLybGlySerPheSerAapAapGluLeuPheIleLeuCyseValGlnProAla 499
Db      1468  ACA-----ACAAGCATGAGTTGTTTGTATGATATGATCCACTT 1509
Qy      500  LysAlaAlaArgLeuLeuThrHisAapLeuArgMetLysThrIlePheLeuArgGlnIle 519
Db      1510  AAACAGTCAAGAAATTAATCTTGCAGTGAATATCCAAACCAAGCAGCACTTTTTCGGAACAA 1569
Qy      520  LeuAapGlnLysAapAapPheProThrAapLysAapCyseLybGlnGlnArgLysCyse 539
Db      1570  TTGATCAAAAGCAGACTTTTCAACAGACAAATGTCAGAAATTAAGAAAGAAAGTGT 1629
Qy      540  GlnAapLeuGlyGlnAapSerLybGluIleThrTrpProCyseHisIleThrLeuGlnGln 559
Db      1630  GAACCTTTAGCGAAGATTCAAATCAAGATTAAGTGGCAGATTCATACGCTAAACAAACAG 1689
Qy      560  CyseAapArgLeuGlyThrThrGlnIleLeuLybGlnValLeuLeuAapGlnHisLysAap 579
Db      1690  TGTATCGCTTGGGAGACTACAGAAATCTTGAACAGGTTTACTAAGATGAACACAGCAT 1749
Qy      580  ThrLeuLysAapGlnLysCyseValLysPheThrLeuLybGluLysCyseAapLysTrpSer 599
Db      1750  ACTTTAAGAACTCATGAAACGTGACAAATATTTAAAGACAAATGTCATAAATGAGTCT 1809
Qy      600  ArgArgGlyAapAapAapPheSerPheValCyseValPheGlnAenAlaThrCyseGluLeu 619
Db      1810  AGAAGGGGTGATGATCGTTCTCTTTGTATGTGTTTACAAAACGCTACGTAAGCTG 1869
Qy      620  MetValLysAapValLysAapArgCyseGlnValPheLybLysAenIleLysAlaSerThr 639
Db      1870  ATAGTAGATGATGTAAGAAAGAGGTGTAAGTATTTGAAAAAATATGCAAGCGTCAAT 1929
Qy      640  IleIleGlnPheLeuGlnAenAenThrAapLysIleThrThrLeuGlnLysArgAapCysePro 659
Db      1930  ATTATATATTTCTTTAAAAAATTAACAAATTAATAACAGATAGCAGCAAAATATTGCTCC 1989
Qy      660  SerTrpHisIleThrLysAapAapArgPheSerProAapCyseProGlyLeuThrLys--Glu 678
Db      1990  TCATGGCACCACATCTCGATAGATTTTACCCATTTCTCTGATCTTTAAAGAAAGAAAA 2049
Qy      679  AapSerCyseThrLysIleLybLysHisArgGlnProPheThrLysArgAlaLeuGln 698
Db      2050  ACTTTCTGTCAAAATCTTTAAAAAATATTGCAACCATCTTCAAAAGAGAAAGTTTAA 2109
Qy      699  AapAlaLeuLysValGlnLeuGlnGlyLysLeuThrAapLysSerLysCyseGlnProAla 718
Db      2110  GATCTCTTAAGTAGAGCTTCAAGGAAATTAAGTAAATTAATTAATGGAATCTGCA 2169
Qy      719  LeuLysArgTrpCyseThrValAlaGlyAapValAenAenAlaSerIleSerGlyLeuCyse 738
Db      2170  TTAGAAAGATTTGCAATATTTGAAAAATGTAAGTATTCATCAATCAACAGTTTATGT 2229

```

```

Qy 739 LysAlaenThrLysAspAsnSerGlySerSerAspGluAspAlaArgLysGluLeuCys 758
Db 2230 AAGATATATACCGAAAGTAAATCTAAAAAGACCGATATAGAAAGTAAAGAAAGACCTTGT 2289
Qy 759 GluLysLeuValLysGluValGluGluGlnCysLysAlaLeuProThrGluLeuGln 778
Db 2290 CTAAATTTAGTGAAGAGGTGAACAGCAATGTAAATGTTCACGACGAAATTTGAGACAT 2349
Qy 779 ProAlaAlaAspLeuLysAspTyrLysThrTyrGluGluLeuLysLysArgAlaGlu 798
Db 2350 GAGGAAAAAGCCTAAAGATGATTTGAAACATTTGAAAACTTAAAAAAACAGCAGAG 2409
Qy 799 GluAlaMetAsnLysSerSerLeuValLeuSerLeuLysLysAsnGluSerVal 818
Db 2410 AAAACATGATTAATTCATCTTTTATCATCTTCCTTAAGAAAGATGAATAATATCA 2469
Qy 819 SerLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGlyLeuGlnAspThr 838
Db 2470 TCGAAAAATAGTACAAAGACAGATAGAAATACCGTTTCAACGAGCTTCAAGATACC 2529
Qy 839 ThrLysHisValLysLysLeuArgGlyValLysAspValSerValThrGluLeuGlu 858
Db 2530 ACAGAACATATGAATAATCTACGAGAGAGATTAAAGATGTATCCGTAAACAGAACTGAA 2589
Qy 859 AlaLysAlaPheAspLeuAlaGluValPheGlyArgTyrValAspLeuLysGluArg 878
Db 2590 GCTAAAGCATTTGATTTGTGTGACAGAAATATTGGAAGATATCTAGACTTAAAGAAAGA 2649
Qy 879 CysAsnLysLeuGlnLysSerAspCysArgLysLysGluAspCysLysAspLeuGluVal 898
Db 2650 TGTAAATTAATGGAATCAGATTTGCGAGATTAAAGAGAGATTCAAGAGATTTGAAGAGAGTA 2709
Qy 899 CysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 918
Db 2710 TGTGAAAGATACCAAGAGATGTTCGAAATTAACCACTGAAGAGTAAACCCGACAGAA 2769
Qy 919 ThrValThrGlnSerThrThrThrThrThrThrThrThrThrThrThrThrThrThr 938
Db 2770 ACAGTACAGAAAGACAAACAGACGACGACGACGACGACGACGACGACGACGACGACGACG 2829
Qy 939 AlaThrGluCysLysSerLeuGlnThrThrAspThrTyrValThrGlnThrSerThrHis 958
Db 2830 GCAACAGATCAATCTTTACAGCAACATCATCATCATCATCATCATCATCATCATCATCAT 2889
Qy 959 ThrSerThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 978
Db 2890 ACCAGCAGCTACATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCA 2949
Qy 979 ArgCysLysPheProThrLysCysThrThrGluGluValAspAspAlaGluAspValLysPro 998
Db 2950 CGTTGCAAAACCAACAGTGTACGACAGG-----GATGATGCGAGAGAGAGTGAAGCCG 3003
Qy 999 SerGluGlyLeuArgMetSerGlyTyrAsnValMetArgGlyValLysValAlaMetVal 1018
Db 3004 AGTAGAGGATTTGAAGATAGTGGGTGAACGATGATGAGGGGGGTGATGTAGCAATGTT 3063
Qy 1019 IleSerPheMetIle 1023
Db 3064 ATTTCGTTTCATGATT 3078

```

RESULT 6

AA294063 standard; DNA; 3042 BP.

AA294063;

15-SEP-2003 (revised)
05-JUN-2000 (first entry)

Pneumocystis carinii major surface glycoprotein gene HMSGp1.
Major surface glycoprotein; MSG; HMSGp1; human; pneumonia; diagnosis; ds.

```

XX OS Pneumocystis carinii; sp. f. hominis.
XX FH Location/Qualifiers
FT CDS
FT 1..3042
FT /tag= a
FT /product= "HMSGp1"
FT /note= "a nucleic acid comprising residues 2894-3042 of
FT this sequence is specifically claimed in Claim 28"
XX
XX W0200009760-A1.
XX
XX 24-FEB-2000.
XX
XX 17-AUG-1999; 99WO-US018750.
XX
XX 17-AUG-1998; 98US-0096805P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Kovacs JA, Huang S, Masur H, Fischer SH, Gill VJ, Mei Q;
XX WPI; 2000-206025/18.
XX DR P-PSDB; AAY79165.
XX
XX Detection of the presence of Pneumocystis carinii in specimens by
XX identification of major surface glycoprotein (MSG) gene sequences using
XX two or more oligonucleotide primers derived from human P. carinii MSG
XX protein encoding sequence.
XX
XX Claim 27; Page 49-53; 110pp; English.
XX
XX The present sequence is that of the novel Pneumocystis carinii sp. f.
XX hominis gene, HMSGp1, which encodes a major surface glycoprotein (MSG)
XX see AAY79165). The gene is 1 of 7 novel, claimed genes (see AA279063-69)
XX of the invention that encode human-P. carinii MSGs (see AAY79165-71). The
XX MSGs include a highly conserved C-terminal region of approximately 100
XX amino acids; this region (also claimed) corresponds to residues 2894-3042
XX of the present sequence. Direct detection or amplification of human-P.
XX carinii MSG-encoding genes, especially by PCR using primers directed at
XX the conserved region of the genes, provides a sensitive and specific
XX technique for the detection of P. carinii, and the diagnosis of P.
XX carinii pneumonia, especially in biological specimens (e.g. blood,
XX sputum) from immunocompromised patients such as those with HIV infection.
XX (Updated on 15-SEP-2003 to standardise OS field)
XX
XX Sequence 3042 BP; 1307 A; 381 C; 606 G; 748 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 3,13e-200 Length: 3042
XX Score: 2762.50 Matches: 541
XX Percent Similarity: 69.10% Conservative: 177
XX Best Local Similarity: 52.07% Mismatches: 280
XX Query Match: 51.05% Indels: 41
XX DB: 3 Gaps: 18
XX
XX US-10-654-416-14 (1-1023) x AA294063 (1-3042)
Qy 1 MetAlaArgAlaValLysArgGlnAlaLysGlyThrGlnAsnSerIleAspGluGluHis 20
Db 1 GTGGCGCGGGGGTTTAAAGCGGAGGTAAACAGACCA---TCAGAGATGATGAGAGAGAA 57
Qy 21 Val-----IleuAlaLeuLysLysGluAspGlyLeuSerGluGlnGluCysLysLys 38
Db 58 GTGGCGCTTTTGGCTTTAATCTAAAGAAAGATTCTAAGATGATTAATAAAATGCGAAGAA 117
Qy 39 LysLeuLysLysTyrCysGlnGluLeuThrGluAlaLysSerAsnIleGluGlnValHis 58
Db 118 AATATTAGAAAAACCTTCGAAAGATTAAGTAAACAATCTAATCCAGAAACAGATCAT 177
Qy 59 ArgLysLeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLys 78
Db 178 GAAGAAGTTAAAGATTTCTGTGATAGCAAAAAACGTGATTAATAAATGTAAAGAACTAAAA 237

```

```
QY 79 AlaasnilegIuLyLeuCySerThrIleuysgIyLysleuLysegluaIlelylys 98
DB 228 AAAAAATGTTGAAAAAATGCGGTGATTTTAAACAAGAAATGGAATGCGTGAAG 297
QY 99 LysIleGIuIleIleThrAspLysAspCysLysgluaengIuGIuCyLeuPheLeu 118
DB 298 GAAGCTTCAAAATTTGAAAAATGATGATACAAATGACAAACAGTGTGTTT 357
QY 119 GluGIyValCySerSerLeuLeuLysAspAspCysAsnThrLeuArgAsnLysCySer 138
DB 358 GAAGAAGCATGCTCT--GATCTTACAAGAATTCACAGATTTTAAGAAACAATGTTAT 414
QY 139 GlnLysLysArgAspLysValAaGIuGIuValLeuLeuArgAlaLeuArg----- 155
DB 415 CAGAAATTAAGCGTGATAGAGTACCAAGAGAGTCTTTTAAGAAATTAATAAGAAAGAA 474
QY 156 --SerAspLeuAsnGlySerValIleCyseGIuLyLysleuLysegluIleCySerProVal 174
DB 475 TTTAAAGATTAATAATTC-----TGTGAAAAATTAACGTGAAGTAACTGTCAAGAA 525
QY 175 MetGIyArgLysSerAspGIuLeuThrAsnLeuCyLeuAsnGlnLysgluThrCySer 194
DB 526 TTAAGTCAAAATGAGCGAATTAATGAAATTAATTTTGAATCAAAAAATACGTGAT 585
QY 195 AsnIleLeuIleGIuLysAspLysLysCyseGIyThr-----LeuLysThrAspVal 211
DB 586 AATCTTGTAAGAAAGAAACCAACCAAAAGTGTGATCTTTCAAAAATCTTAAACCGAAAT 645
QY 212 SerAlaIleLeuGIySerPheLys--LysGIuThrCyLeuGIuLeuGIuGIuGIu 230
DB 646 AAAAAATTAAGAAAGATGACACTTAAAAAATGCCCCATTTTATTAAGAAAGATC 705
QY 231 TyrPheTyrlleGIyAsnCyseGIyAspAspAspIleIleLysCyseIleGIuLeuGIy 250
DB 706 ATTTTATTAATGAAGAAATGTTGTGAACGATTC---CTGAAGTGAAGTGAATGGA 762
QY 251 LysCyseGIuGIuGIuAsnIleAlaTyMetProGIyProAspPheAspProThrArg 270
DB 763 AAATGTCAAGAAAGAAATATTAATTAATTAATTAATTAATTAATTAATTAATTA 822
QY 271 ProGIuAlaThrIleAlaGIuAspIleGIyLeuGIuGIuPheTyrlLysValGIuGIu 290
DB 823 CCAGAAATTAATCAATGACAGAAAGATTAATTAATTAATTAATTAATTAATTA 882
QY 291 AspGIyValPheIleGIyLysAsnHisLeuArgAspAlaThrAlaLeuLeuLeu 310
DB 883 GAAGAACTCTGTTGGGAAACCTTACCAAGCATGCTACTTGTGTGCAATTTTG 942
QY 311 IleGlnAspSerSerLeu--LysLysLysAspAspLysGIuLysCyseGIuGIuAlaLeu 329
DB 943 ATTCAAGATCCATCTCTTAACAACCAAGCACTTAACAAAGAAATTAATTAATTTCT 1002
QY 330 GlnLysSerCyseLysAsnProHisGIuHisAlaLeuGIuSerLeuCyseLysAsn 349
DB 1003 GAAGTAAATGTAATAATTAATAAGAACTGATTAATTAATAAGTCTATGCGAGATTA 1062
QY 350 GlnLysSerAsnAspGIyThrLysLysCyseGIuGIuLeuGlnAsnAspIleAsnLysThr 369
DB 1063 AATGCAATTAATAAGATTAAGCAAAATATCGAAGAACTTAAGTACGATTTGA 1122
QY 370 CyseLysIlePheThrSerLysValThrAsnAsnArgLeuPheAspProThrLysGIyAsn 389
DB 1123 TGTAAATTTTCAATTCAAAAACCTTATGATTCATTTTGGCGATGAAATTAAT 1182
QY 390 AsnGIuIleValGIyTrGIuGIuLysLeuProThrPheLeuSerAsnGlnAspCyseValLys 409
DB 1183 GATGCAATTAATAATTAAGGGAATTAATCAACGTTTCTTAAGCAATTAAGATTTGA 1242
QY 410 LeuGIuSerTyrlCysePheTyrlPheGIuLysLysCyseProAspGIyGlnAsnAlaCyseLys 429
DB 1243 TTAGATGCTATGTCTTATTTTGAAGAAAGCTGT--AGAAGGAAAGCTGCAAG 1299
QY 430 AsnIleArgAlaThrCyseTyrlLysArgGIuLeuAspAlaArgAlaAsnLysValLeuGln 449
DB 1300 AATATCAGAGCAGCATCTCTCAAGAGAGACTTGAACATTAAGAAATGAAGTATTA 1359
QY 450 GlnAsnMetArgGIyMetLeuIleGIySerAsnLysSerTrpLeuGIuLysPheGIuGln 469
DB 1360 AAAGAAATGCGAGAAATGCTGCAATGTTCAAAATTAACATGCTTAAGTGTTC 1419
QY 470 GlnLeuLysValCyseGIuLysLeuLysGIuAsnLysGIySerPheSerAsnAsp 489
DB 1420 AAACCTAGAAAGTGTCAAAAAAGTGAATAAGAAATTAAGAGATTTTCCAGTAA 1479
QY 490 GlnLeuPheIleLeuCyseValGlnProAlaLysAlaIleArgLeuLeuThrHisAspLeu 509
DB 1480 GAATTAATTTGCTTATGTGTAACAACATCAAAACAGCTGATGCTTTCGATGATCTT 1539
QY 510 ArgMetLysThrIlePheLeuArgGIuLysAspGIuLysAspGIuLysAspPheProThrAsp 529
DB 1540 CGGATGAAATTAATCTTTTGCAGAGCATTTGAACAGAAAGCGAGATTTCCAGTGA 1599
QY 530 LysAsnCyseLysGIuLeuGIyArgLysCyseGIuAspLeuGIuAspSerLysGIuIle 549
DB 1600 GAAGACTGCGAAGATTAATTAAGAAATGTGAGGCTTAAGAAAGATTTCAAAAAAT 1659
QY 550 ThrTrpProCysHisThrLeuGIuGIuGlnCysAsnArgLeuGIyThrThrGIuIleLeu 569
DB 1660 GAATGGCCATGTCTATACATTAAGCCAAATTTGATATCAATTAAGAAACGCTAAAGAA 1719
QY 570 LysGlnValLeuLeuAspGIuHisIleLysAspThrLeuLysAspGIuGIuSerCyseValLys 589
DB 1720 AAAGAACTTTTACATTAATTAAGAAATTAAGATTAATTAATTAATTAATTAATTA 1779
QY 590 TyrlLeuLysGIuLysCyseAsnLysTrpSerArgArgLysAspAspPheSerPheVal 609
DB 1780 TATTTGAAGAGAAATCAATGAATGCTTGAAGAGAAATGAACGTTCTCTTTT 1839
QY 610 CyseValPheGlnAsnAlaThrCyseGIuLeuMetValLysAspValLysAspArgCyseGIu 629
DB 1840 GTGCTTTGCAAAATGAGACTTGAAGATTAATGTAAGTGAAGTGAAGTGAAGTGA 1899
QY 630 ValPheLysLysAsnIleLysAlaSerTyrlleGIuPheLeuGlnAsnThrAsn 649
DB 1900 ATATTTGAAAAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTA 1947
QY 650 LysIleThrThrLeuGIuArgAsnCyseProSerTrpHisThrTyrlCyseAsnArgPheSer 669
DB 1948 AAATTAAGAAATTAAGCAATATTTGTCCTATTTGGCACCACTGCAATGAATTTG 2007
QY 670 ProAsnCyseProGIyLeuThrLysGIuAsnSerCyseThrLysIleLysLysHisArgGIu 689
DB 2008 CCCAATTTGCCCGATCTTGAATA--ATAAATGTGAAGCTTTGAATAATTTGCA 2064
QY 690 ProPheTyrlLysArgLysAlaLeuGIuAspAlaLeuLysValGIuLeuGIuGIyLysLeu 709
DB 2065 CTTATTTTAAGCAAAAGACCTTGAATTAATTAATTAATTAATTAATTAATTAATTA 2124
QY 710 ThrAspLysSerLysCyseGIuProAlaLeuLysArgTyrlCyseThrValAlaGIyAsnVal 729
DB 2125 GATTAAGAAAAAACTGCAAAACAATCTTGAATTAAGTACTTAACATTAATGATCA 2184
QY 730 AsnAsnAlaSerIleSerGIyLeuCyseLysAlaAsnThrLysAspAsnSerGIyLysSer 749
DB 2185 GGAATTAATAACATTAAGGTTTGTGAACGTTCTACTGATTAACAT----- 2232
QY 750 AspGIuAspAlaArgLysGIuLeuCyseGIuLysLeuValLysGIuValGIuGIuGIu 769
DB 2233 --GAAACATTTAAGATTAATCTTTGCAAAATTAATTAATTAATTAATTAATTA 2289
QY 770 LysAlaLeuProThrGIuGIuGIuGIuProAlaAlaAspLeuLysLysAspTyrlLysThr 789
DB 2290 CAAGGATTAATCAAAAGAACTGGAAGAAAGCAAAATATGTTTAAGAAAGAAACATTA 2349
QY 790 TyrlGIuGIuLeuLysLysArgAlaGIuGIuAlaMetAsnLysSerSerLeuValLeuSer 809
```


Qy	875	LeuLYSGVAGSYASNTLyeucluserApsCYsargIleLYSGluAspCYeLYsAsp	894
Db	2557	TTGAAAGAAATATGTATTCATTCTACTGTAAGAATGTGGTTCCAAAAGAGAGT---GAC	2613
Qy	895	LeuGLUGluValCYeLYsLYsIleasnLYsAlaCYsArgAsnLeuLYsProLeuGLuVal	914
Db	2614	TGTAGGATTCATGTATTAAGATACAGGGAATATGTTCAACTTAGAGCACTAAAGTG	2672
Qy	915	LYsProHISGLuThrValThIGlu-----SerThrThrThrThrThrThrThrThr	932
Db	2674	AGACCAACCGAAATAGTAACCTAAAAACATACACTAACCAACCAACCAACCACT	2733
Qy	933	ThrValAlaAspProLYsAlaThrGLUCysLYsSerLeuGINThrThAspThrTryAl	952
Db	2734	ACCAATTAAAGCGCAAGGCAACAGACTGCCACTTTTACAGACAACAGTACGGGTC	2793
Qy	953	ThrGINThSerThrHISThrSerThSerThrIleThSerThrIleThSerLYsIle	972
Db	2794	ACAAAGACGTGCACCCTACTAGCAATCCACACCAATCATCTACAGTCAACGTCAGAA	285
Qy	973	ThreuthrSerThrArgArgCYeLYsProThrLYsCYeThrThGLYGLuAspAsp	992
Db	2854	ACGTTGACCTGCACAAAGACGGGTGAAGCTTACGAAGTGTACACAGAGAGAGATGAA	2913
Qy	993	AlaGLYAspValLYsProSerGLUGlyLeuArgMetSerGLYTPAsnValMetArgGly	1012
Db	2914	GCAGAGACGTGAAACCAAGCTGAAGGTTGAGATGATGGATGGAGTGTATGAGGGGG	2972
Qy	1013	ValIleValAlaMetValIleSerPheMetIle	1023
Db	2974	GTTGTTATTAGCAATGACGATTTTCATTCAATGATT	3006
RESULT 8			
AAQ41226			
ID	AAQ41226	standard; DNA; 3521 BP.	
XX	XX		
AC	AAQ41226;		
XX	XX		
DT	27-AUG-2003 (revised)		
DT	17-DEC-2001 (revised)		
DT	02-SEP-1993 (first entry)		
XX	XX		
DE	Clone GP3 encoding major surface gp of rat P. carinii.		
XX	XX		
KM	Major surface glycoprotein; gp16; rat; Pneumocystis carinii; vaccine;		
KM	HIV; human immunodeficiency virus; diagnostic; PCR; ss.		
XX	XX		
OS	Pneumocystis carinii.		
XX	XX		
PN	USN7958683-N.		
PD	01-APR-1993.		
XX	XX		
PF	09-OCT-1992; 92US-00958683.		
XX	XX		
PR	09-OCT-1992; 92US-00958683.		
XX	XX		
PA	(USSH) US DEPT HEALTH & HUMAN SERVICE.		
XX	XX		
PI	Kovace JA, Angus CW, Powell F, Edman JC;		
XX	XX		
DR	WPI; 1993-159487/19.		
XX	XX		
PT	Major surface glyco-protein of Pneumocystis carinii, encoded by multi-		
PT	copy gene family - used in a vaccine and diagnostic assay for P. carinii		
PT	infection in AIDS patients.		
XX	XX		
PS	Disclosure; Page 37; 50pp; English.		
CC	Multiple clones were identified by immunoscreening a rat P. carinii cDNA		
CC	library using rat serum generated against whole rat P. carinii. Clones		
CC	reactive with polyclonal serum were evaluated to identify those encoding		

CC p116, three such clones (PC3, PC5 and PC14) were sequenced and contained
CC ORFs encoding closely related but distinct proteins. Although none of the
CC clones contained the complete coding sequence, overlapping regions
CC allowed alignment of the three clones and generation of a putative
CC composite sequence encoding a protein of ca. 122 kD. PC5 was used to
CC screen a second cDNA library constructed in modified lambda ZAP, YcBcl1.
CC Four positive clones, GP3, GP22, GP46 and GP14 were sequenced and were
CC found to contain ORFs encoding proteins similar to the original gp116
CC clones. The gp116 of P. carinii can be used in a vaccine against
CC infection in HIV-infected individuals and also as a diagnostic agent.
CC Study of the expression of the protein should lead to understanding its
CC role in the pathogenesis of P. carinii pneumonia and may lead to new
CC strategies for control or prevention of infection. The DNA sequence may
CC be used to make PCR primers for diagnostic use. NB. Due to a poorly
CC reproduced sequence several unreadable bases are represented by N in the
CC sequence below. See also AAQ41223-39. (Note: Revised entry submitted to
CC correct the patent number format of US Government-owned NITs applications
CC to prevent clashes with ongoing US granted patent numbers. For further
CC information please visit the Derwent web site at
CC www.derwent.com/dmpi/updates/ntis_us.html.) (updated on 27-AUG-2003 to
CC correct OS field.)

Alignment Scores:	
Pred. NO.:	9 328-99
Score:	1430.00
Percent Similarity:	50.53%
Best Local Similarity:	33.30%
Query Match:	26.43%
DB:	2
	Gaps: 28
	Length: 3521
	Matches: 346
	Conservative: 179
	Mismatch: 360
	Indels: 154
	Gaps: 28

US-10-654-416-14 (1-1023) X AAQ41226 (1-3521)

Qy	96	ILeYsrlYsrlIeGlnIllelthraspYsaspCySlySGlYsnglGlnGncs	115
Db	422	CTTCAAAAAGTATTGAATTAATATAAAGATATAAATTGGCAAAAATATGAAGAAAATGT	481
Qy	116	LeuPheLeuGluGluValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsn	135
Db	482	ATACTTTTAGAAGAGCAGGATTATGATGTTATTAAAGGTAATCTGATTCGAATTGAGGAA	541
Qy	136	LysCysTrgTglnYslySargAspLysValAlaGluGluValLeuLeuArgAlaLeuArg	155
Db	542	GGAGTGTACAAATTTGAAGCGTGAAGAAAGGTGCAGAGGAGCTTCTTGAGGGCGCTCCGA	601
Qy	156	SerAspLeuAsnGlySerValIleCysGluLysLysLeuYsGluIleCysProValMet	175
Db	602	GGGAGTCTTAAAGAAGACTTAAATGTAAAGAAAGATGAATCTGTTGCCAGTGTGG	661
Qy	176	GlyTrgGluSerAspGluLeuThrAsnLeuCysLeuAsnGlnYsGluThrCysLysAsn	195
Db	662	AGCCGAGAAACCGACGAATTGATGTCTTTTGCCCTTGATTTCTGCTAAACATGTGCAAT	721
Qy	196	IleLeuIleGluYsAspLysLysCysGlyThrLeuYsThrAspValSerAlaIleu	215
Db	722	CTG-----AAAAAAAATTTGGGTACTGTTTGGCGAGCCTTTAAAAAAGAGCTT	769
Qy	216	GlySerPheLys---LysGluThrCysLeuGluLeuLeuGlnGlnCysTyPheThrIle	234
Db	770	AAAGATACCAAGATTAGCGGAAAGAGTGCATATAAGACTTGAAATGTCATTTTTTAAACGA	829
Qy	235	GlyAsnCysGlyAspAspAspIleIleLysCysIleGluLeuGlyGlyLysCysGlnGlu	254
Db	830	GAACCGTGTATGAT-----GCGAAATGCAAGAAGTTTAGAGAGCATGCAAGGA	880
Qy	255	GlnAsnIleAlaTyMetProProGlyProAspPheAspProThrArgProGluAlaThr	274
Db	881	AAAATATTTATATATTAAGCGCCAGCAATCTGATCTTGCTTCGACAGCCAGGCGCTCC	940
Qy	275	IleAlaGluAspIleGlyLeuGluGluLysPheTyTrpLysValGluGluAspGlyValPhe	294
Db	941	TTGTTGAGAACTATTTGGTGGATGATGTGTATAAAAACCGGAAAAACATGAGGATTATT	1000


```
Db 3053 GAAACAAAGCTGATGCGACGCTGATGAAAGACCGTTGAGAGACTGTACAGAAACC 3112
Qy 941 -----GluCysLeuGlnThrAspThr 950
Db 3113 AAGTCACTAGGTGAGAGAAAGTAAACAGAGAGTCAATGATCAAAACAGATGACA 3172
Qy 951 TrpValThrGlnThrSerThrHisThrSerThrSerThrIleThrSerThrIleThrSer 970
Db 3173 TGGGTGACGACTGACTGATTCATGACAGTACGACACAGAGTACGACCGTACGCTCG 3232
Qy 971 LysIleThrLeuThrSerThrHisThrArgArgCysLeuProThrIleCysLeuThrGly----- 988
Db 3233 ACAGTACGCTGACTGCTGATCGACGACGCTACCAATGATGACACCGATTCAGAC 3292
Qy 989 -----GluGluAspAspAlaGlyAspValLysProSerGluGlyLeuArgMet 1004
Db 3293 AAAGAGACACAGAAAGAAAGATGATGAGAGTGAACCGAATGAGGGAAATGAAATA 3352
Qy 1005 SerGlyTrpAsnValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023
Db 3353 AGAGTCTGATGATGATTAATAATATGTTGCGAGTGTGTTATGGGATGATG 3409

RESULT 9
AAQ41230
ID AAQ41230 standard; DNA; 3521 BP.
XX
AC AAQ41230;
XX
DT 27-AUG-2003 (revised)
DT 17-DEC-2001 (revised)
DT 02-SEP-1993 (first entry)
XX
DE Gene encoding major surface GP of rat P. carinii.
XX
KW Major surface glycoprotein; gp16; rat; Pneumocystis carinii; vaccine;
KW HIV; human immunodeficiency virus; diagnostic; PCR; ss.
XX
OS Pneumocystis carinii.
XX
FH Key Location/Qualifiers
FT misc_feature 1..722
FT /*tag= b
FT /note= "fragment determined by PCR"
FT CDS 146..3412
FT /*tag= a
FT misc_feature 626..3521
FT /*tag= c
FT /note= "fragment from original GP3 clone"
XX
PN USN7958683-N.
XX
PD 01-APR-1993.
XX
PF 09-OCT-1992; 92US-00958683.
XX
PR 09-OCT-1992; 92US-00958683.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Kovace JA, Angus CW, Powell F, Edman JC;
DR WP1: 1993-159487/19.
DR P-PSDB; AAR36706.
XX
XX Major surface glyco-protein of Pneumocystis carinii, encoded by multi-
XX copy gene family - used in a vaccine and diagnostic assay for P. carinii
XX infection in AIDS patients.
XX
XX Disclosure, Page 41; 50pp; English.
XX
CC Multiple clones were identified by immunoscreening a rat P. carinii cDNA
CC library using rat serum generated against whole rat P. carinii. Clones
```

```
CC reactive with polyclonal serum were evaluated to identify those encoding
CC gp16. Three such clones (PC3, PC5 and PC14) were sequenced and contained
CC ORFs encoding closely related but distinct proteins. Although none of the
CC clones contained the complete coding sequence, overlapping regions
CC allowed alignment of the three clones and generation of a putative
CC composite sequence encoding a protein of ca. 122 kD. PCR was utilized to
CC determine the full sequence of the gene. The 5' end of the message was
CC identified by anchored PCR using primer JK58, which spanned the putative
CC start codon of the composite protein. The intervening region was
CC determined by reverse transcription followed by PCR using primers
CC spanning the 5' end to base 722 in GP3. A single clone was identified
CC that had an identical sequence to the first 76 bases of GP3. The
CC complete, composite cDNA contained an ORF encoding a protein of ca. 122
CC kD. The gp16 of P. carinii can be used in a vaccine against infection in
CC HIV-infected individuals and also as a diagnostic agent. NB. Due to a
CC poorly reproduced sequence several unreadable bases are represented by N
CC in the sequence below. See also AAQ41223-39. (Note: Revised entry
CC submitted to correct the patent number format of US Government-owned NTIS
CC applications to prevent clashes with ongoing US granted patent numbers.
CC For further information please visit the Derwent web site at
CC www.derwent.com/dwpl/updates/ntis_us.html.) (Updated on 27-AUG-2003 to
CC correct OS field.)
XX
SQ Sequence 3521 BP; 1238 A; 385 C; 695 G; 725 T; 0 U; 478 Other;

Alignment Scores:
Pred. No.: 2,06e-70 Length: 3521
Score: 1057.50 Matches: 333
Percent Similarity: 43.21% Conservative: 151
Best Local Similarity: 29.73% Mismatches: 469
Query Match: 19.54% Indels: 170
DB: 2 Gaps: 32

US-10-654-416-14 (1-1023) x AAQ41230 (1-3521)
Qy 1 MetAlaArgAlaValLysArgGlnAlaIle---GlyThrGlnSerIleAspGlu 19
Db 146 ATGGCAGCGCCGGTTAGAGGACAGCAGTACAGAGCACAGATGATGATGAGAA 205
Qy 20 HisValLeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGlnGluCysLysLys 39
Db 206 CANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGAAACAAATAATGCAAGAA 265
Qy 40 LeuLysLysThrCysGlnGluLeuThrGluAlaLysLeuAsnIleGluGlnValHisArg 59
Db 266 CTCGAGAAATATTGTAAAGAGTTGAAGAGACCATTAATAATGTAGCAATGTGATGAT 325
Qy 60 Lys-LeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAl 79
Db 326 AAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGAAAAAATGCAAGACGTGAAAA 385
Qy 79 AsnIleGluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLys 99
Db 386 AAAAGTTGAAGATGATTAATA-----GATTTTGAAGAGGAACTTCAAAAAGT 433
Qy 99 sIleGlnIleIleThrAspLysAspCysLysGluAsnGluGlnCysLeuPheLeuG 119
Db 434 ATTGAATTAATTAATAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTACTTTTGA 492
Qy 119 uGlyValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyrG 139
Db 493 AGAGAGCGATTGATGATTTTAAGATTAAGTATGATGATGAGGAGAGATGTTCAA 552
Qy 139 nLysLysArgAspLysValAlaGluGluValIleLeuArgAlaLeuArgSerAspLeu 159
Db 553 ATTGAACCGTGAAGAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGGGGTGCTAA 612
Qy 159 nGlySerValIleCysGluLysLysLeuLysGluIleCysProValMetGlyArgG 179
Db 613 AGAAGAACTTAATGTAAGGAAAGATGATATCTGTGCGCATGTGAGCCGAGAAAG 672
Qy 179 rAspGluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIleLeuIleG 199
```


Db	1131	GTTAAGAGCACAACAAGTAGTACAGTACACAAGTAAAGTAAAGAGAACACCTT	1190
Qy	22	LeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGluGluCysLysLysLysLys	41
Db	1191	TTGGCTTTCATTGTGAAGAGCAAAACATGATGATGAGTAAGTAAATCAAAAAAGGCTCGAG	1250
Qy	42	LysTyrCysGluGluLeuThrGluAlaLysLeuIleGluGluValHisArgLysLeu	61
Db	1251	GAATATGTGTAAAGAGTTGAAGAAAGCAAGTACAGAAATTTCC--AGTGTGAATGAGAAAGTT	1307
Qy	62	LysGluPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAlaAsnIle	81
Db	1308	AAAGACCTTTGTGATGATATAAAAAAGAGACCAAAAAATCAAGAACTGAAAAAAAAGTT	1367
Qy	82	GluLysLysCysThrThrIleLysGlyLysLysGluAlaIleLysLysLysIleGlu	101
Db	1368	AAAGATGAATTCGGAACTTTTGATACCGAGTCTTGAAAGCATCGTAGAT-----GAC	1418
Qy	102	IleIleThrAspLysAspCysLysGluAsnGluGluGluGluCysLeuPheLeuGluLysAl	121
Db	1419	ATAGAAAGATGAGAGAGTTGTAAAAAACAATGAAAGAAATATATACTTTTGAGAGAAACA	1478
Qy	122	CysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyrGluLysLys	141
Db	1479	GACCAAAATAGCTTTAAGAGAACTGTCTCAAGTTGAGGAAAGAGTGTACAAATTGAG	1538
Qy	142	ArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGlySer	161
Db	1539	CGTAAAAAGGTCGACGAGAGAGCTCTTTTGAGGGGGCTCGAGGGGAGATGCTAAAGATGA	1598
Qy	162	ValIleCysGluLysLysLysLysGluIleCysProValMetGlyArgGluSerAspGlu	181
Db	1599	GCTAAATGTAAAGAAAGATGAAACTGTTCGCCAAATGTTAAAGCCGGAAGAGTGAAG	1658
Qy	182	LeuThrAsnLeuCysLeuAsnGluLysGluThrCysLysAsnIleLeuIleGluLysAsp	201
Db	1659	CTGATGTTTCTTCTGCTTGAATTCGGATGAACGTTAAAGCCGTGAACAAATTCAGAA	1718
Qy	202	LysLysCysGlyThrLeuLysThrAspValSerAlaIleGluSerPheLysLysGlu	221
Db	1719	GAGGTTTCCCGCTTTAAAAAGAAAGCTTAA-----GATCGCGAATTTAAAGGAA	1769
Qy	222	ThrCysLeuGluLeuLeuGluGluGluCysTyrPheTyrIleGlyAsnCysGlyAspAsp	241
Db	1770	AAATGCTATGAAAGACTTGAGAAATGTCATTTTAAAGAAAGCGCTGATCGAA-----	1823
Qy	242	IleIleLysCysIleGluLeuGlyLysCysGluGluGluAsnIleAlaTyrMetPro	261
Db	1824	--ACAAAGTGATGATGAGATATGAGCAAGCAAGCAAGAAAGAAAGATTCACTATTAAGCG	1880
Qy	262	ProGlyProAspPheAspProThrArgProGluAlaThrIleAlaGluAspIleGlyLeu	281
Db	1881	CCGGAATCTGATTTTAAGTCCTGTCAAGCGCAAGCGCTGTTTGAGAAAGTAATTCGGTTG	1940
Qy	282	GluGluPheTyrLysLysValGluGluAspGlyValPheIleGlyLysAsnIleLeu--	300
Db	1941	GATGATGTGTATAAAAAGCTGAAAAAAGAAAGAAATTATTATTTGGAATAATCAGAGTGGAT	2000
Qy	301	-----ArgAspAlaThrAlaLeuLeuAlaLeuLeuIleGluAspSerSerLeuLys	317
Db	2001	CTACCAAGAAAGTCAGTACAAATTTTGGCAAGTCTCTTGACTGATCTGTGAGAGAGAT	2060
Qy	318	LysLysAspAspLysGluLysCysGluGluGluAlaLeuGluLysSerCysLysAsnProHis	337
Db	2061	GAGATGATGACGAGGAAGAAATTCGGTAAAGCGTTTAAAGAAATGTGAACCTTTAAGAT	2120
Qy	338	GluHisGluAlaLeuGluSerLeuCysLysLysAsnGlyLeuSerAspAspGlyThrLys	357
Db	2121	TTGAATACGATTTTATGAGAGTTATGCAAA-----GATGCTGATTAAGAAAT	2168
Qy	358	LysCysGluGluLeuGluAsnAspIleAsnLysThrCysLysIlePheThrSerLysVal	377
Db	2169	AAATGCAAAAA--AACCTAGTGTAAAGAAAGATGTT-----	2204

OY	378	Thrsbnaabnrqleupheapbprcfrthlysgly-----	388
Db	2205	ACAAAACCTCAAGTTAAATCTTTATGTGCAAAAGGGGTGTCTACGAGTTTAAAGAACTAAA	2264
OY	389	AenabngluilevialgyltrpgluuglyleuprothrpheuseraenngluabpCybAla	408
Db	2265	AAATCAACATCTTTATGTGTGGGACAGCTTCCAAACATATTATTCGAAGGAGAGTGTGCA	2324
OY	409	LysleugluserTyrcybpheTyrpheglulyslyscyspbaapglgu-----Aen	426
Db	2325	GAACCTGAGTCGCAATGTTCTATTATTTGAAAAAGCGGTAAAGATATATGATTTGGTGA	2384
OY	427	AlaCyslysbaniIeaagalaIatnCybTyrlvlybArglyleuabpAlaagAlaenlysc	446
Db	2385	GGCGTCAAAATCTACATCATGACGGTGTATTAATAAGGACACAGATGTTGAATAG	2444
OY	447	ValleuglngluabmetcArglymetleu---HisglyserabnlyssertPleuglu	465
Db	2445	TTCTTTCAAAAGAAATTGAAAGGAAAGCTTGGTCATGTATAGATTATTAACGATCCATA	2504
OY	466	LysbhegnglnglualeuVallybValCysegulysleulyslysgluabnlysglyser	485
Db	2505	GATTGTAAAAAATATGTGTGTAAGAAACGTGTACAAAACCTTAAAAAGATATAAAGA-----	2558
OY	486	PheserabnapgluabpheiIeLeuCybValGlnProhlaIybaIaIaIaargleuLeu	505
Db	2559	-----TACCTTTCAAAATAGCTTTATCTTAAGAAACATGTATTGGCTT	2603
OY	506	ThrlsbapleuabrgmetlysrthrllePheleuabrglnglnleuabpGlnlybArGasp	525
Db	2604	TCAATATGATATTTTCTCCATCCAAAGAGTTAAGTTCGCTTTAATGATGATCAGAGAT	2663
OY	526	PheprothrlepylysbancylbysglulnleuglyArglyscylnglnleuglyguabp	545
Db	2664	TTTCCATTTGAAAAGATTGCTTGTAATTGGAGAGAGATGATCAACTTAAGTAGAT	2723
OY	546	SerlysgluilethrTrpProCybHisThrleuglnglnglnCysbaenArgleuglyThr	565
Db	2724	TCAATTATTTGAATTGAAAAAGTGTATATACATTGAAAAAGACCGTGTGATATTTTGACGT	2783
OY	566	ThrgluileLeuylsglnValleuLeu	574
Db	2784	ACAGAAAGATTTCAGAAAAGTATTTTA	2810
RESULT 14			
AAQ41228			
ID	AAQ41228	standard; DNA; 2110 BP.	
XX	AAQ41228;		
AC			
XX			
DT	27-AUG-2003	(revised)	
DT	17-DEC-2001	(revised)	
DT	02-SEP-1993	(first entry)	
XX			
DE	Clone GP22 encoding major surface gp of rat P. carini.		
XX			
KM	Major surface glycoprotein; gp16; rat; Pneumocystis carinii; vaccine;		
XX	HIV; human immunodeficiency virus; diagnostic; PCR; ss.		
OS	Pneumocystis carinii.		
XX			
PN	USN7958683-N.		
PD	01-APR-1993.		
PF	09-OCT-1992;	92US-00958683.	
XX			
PR	09-OCT-1992;	92US-00958683.	
XX			
PA	(USSH) US DEPT HEALTH & HUMAN SERVICE.		
XX			
PI	Kovacs JA, Angus CW, Powell F, Edman JC;		


```

QY 866 ALaGluValPheGlyArgTyrValAlaSerLeuValGluArgCysAsnLysLeuGluSerAsp 885
Db 1489 ACAATAGCATTTGGAGTTGATTGTAATTTGAAAGGAAATGCGCTTTCACACTAGAT 1548
QY 886 CysArGlieLysGluAspCysLysAspLeuGluValCysLysLysIleAsnLysAla 905
Db 1549 TGGCGTTTAAAGAAAGATTTGTCAGTGTTCGAGTGTTCGAAAGAAATAGACAAGTTA 1608
QY 906 CysArGlnLeuLysPheLeuGluValLysPheHisGluThr----- 919
Db 1609 TGT---GAAGTCGAAACCATTAAGATTACGCTCATCATACAGACAAATTAACAATAG 1665
QY 920 ValThrGluSerThrThrThrThrThrThrThrThrThr-----ValAlaAspPro 937
Db 1666 GTGACGGAAACGAAACGCGAAACAAAGACAAAGAACTGATGACAAAGCTGATGAG 1725
QY 938 Lys-----AlaThr 940
Db 1726 AAGACCGGTACGAAACCTTTACAGAAACCAAGCAATAGGTGAGGAAAGTAAACAGAA 1785
QY 941 GluCysLysSerLeuGlnThrThrAspThrTyrValThrGlnThrSerThrHisThrSer 960
Db 1786 GAGGTGACAAATGTTCAAAACAAAGATACATGATTAACAGTACGTCATTCATCGAGT 1845
QY 961 ThrSerThrIleThrSerThrIleThrSerLysIleThrLeuThrSerThrArgArgCys 980
Db 1846 ACGACAAACGACACCTCAACGCTGACGTCACAGTACGACCTGACCTGATCGGCAAGTGC 1905
QY 981 LysProThrLysCysThrThrGly-----GluGluAspAspAlaGly 994
Db 1906 AAGCTTACCAATGTATCCACATGATTCACAAAGACACAGAAAGAAAGATGATGAA 1965
QY 995 AspValLysProSerGluGlyLeuArgMetSerGlyTyrAsnValMetArgGlyValIle 1014
Db 1966 GAAGTGAACCGAATGAGAGGAATGAATTAAGATTCTTGATATGATTAATAATATGTTG 2025
QY 1015 ValAlaMetValIleSerPheMetIle 1023
Db 2026 TTGGAGTGAATTTGATCGGGAGTATG 2052

RESULT 15
AAQ41227
ID AAQ41227 standard; DNA; 2058 BP.
XX
AC AAQ41227;
XX
AC 27-AUG-2003 (revised)
DT 17-DEC-2001 (revised)
DT 02-SEP-1993 (first entry)
XX
DE Clone GP46 encoding major surface gp of rat P. carinii.
XX
KW Major surface glycoprotein; gp116; rat; Pneumocystis carinii; vaccine;
KW HIV; human immunodeficiency virus; diagnostic; PCR; ss.
XX
OS Pneumocystis carinii.
XX
PN USN7958683-N.
XX
PD 01-APR-1993.
XX
PD 09-OCT-1992; 92US-00958683.
XX
PR 09-OCT-1992; 92US-00958683.
XX
PR 09-OCT-1992; 92US-00958683.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Kovacs JA, Angus CW, Powell F, Edman JC;
XX
DR WPI; 1993-159487/19.
XX
PT Major surface glyco-protein of Pneumocystis carinii, encoded by multi-
copy gene family - used in a vaccine and diagnostic assay for p. carinii

```

```

PT infection in AIDS patients.
XX
XX Disclosure; Page 3q; 50pp; English.
XX
CC Multiple clones were identified by immunoscreening a rat P. carinii cDNA
CC library using rat serum generated against whole rat P. carinii. Clones
CC reactive with polyclonal serum were evaluated to identify those encoding
CC gp116. Three such clones (PC3, PC5 and PC14) were sequenced and contained
CC ORFs encoding closely related but distinct proteins. Although none of the
CC clones contained the complete coding sequence, overlapping regions
CC allowed alignment of the three clones and generation of a putative
CC composite sequence encoding a protein of ca. 122 kD. PC5 was used to
CC screen a second cDNA library constructed in modified lambda ZAP. YcDe11.
CC Four positive clones, GP3, GP22, GP46 and GP14 were sequenced and were
CC found to contain ORFs encoding proteins similar to the original gp116
CC clones. The gp116 of P. carinii can be used in a vaccine against
CC infection in HIV-infected individuals and also as a diagnostic agent.
CC Study of the expression of the protein should lead to understanding its
CC role in the pathogenesis of P. carinii pneumonia and may lead to new
CC strategies for control or prevention of infection. The DNA sequence may
CC be used to make PCR primers for diagnostic use. See also AAQ41223-39.
CC (Note: Revised entry submitted to correct the patent number format of US
CC Government-owned NTRIS applications to prevent clashes with ongoing US
CC granted patent numbers. For further information please visit the Derwent
CC web site at www.derwent.com/dwpi/updates/ntris\_us.html.) (Updated on 27-
CC AUC-2003 to correct OS field.)
XX
SQ Sequence 2058 BP; 821 A; 290 C; 472 G; 475 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 1,34e-55 Length: 2058
Score: 859.00 Matches: 209
Percent Similarity: 49.37% Conservative: 106
Best Local Similarity: 32.76% Mismatches: 225
Query Match: 15.88% Indels: 98
DB: 2 Gaps: 15

US-10-654-416-14 (1-1023) x AAQ41227 (1-2058)
QY 465 GluLysPheGlnGlnGluLeuValValCysGluLysLeuLysGluAsnLysGly 484
Db 70 GAGGAATGCCAAATAATCTGTGTAAGAGCTGTAAATACTT-----GAATAATAA--- 120
QY 485 SerPheSerAsnAspGluLeuPheIle---LeuCysValGlnProAlaLysAlaAlaArg 503
Db 121 -----GATTAAGATATACCTTCCAAGAGTGTCTTATCTTAAGAACTATGTTAT 168
QY 504 LeuLeuThrHisAspLeuArgMetLysThrIlePheLeuArgGlnGlnLeuAspGlnLys 523
Db 169 ATGCTTTCAGATGATATATTTCTTCAATCCAAAGAGTTGGAGCGCTTTGGATGATCAA 228
QY 524 ArgAspPheProThrAspLysAsnCysLysGluLeuGlyArgLysCysGlnAspLeuGly 543
Db 229 ACGGATTTTTCATTAAGAAAGCATTTGTGTAATTAAGAGAGAGTGTGATGAACCTGAA 288
QY 544 GluAspSerLysGluIleThrTyrProCysHisThrLeuGlnGlnGlnCysAsnArgLeu 563
Db 289 ACTTATTCACATTTGGAATTCGGAAGAGTATTAACATTTGAGAACCGCTGTGAATACCTT 348
QY 564 GlyThrThrGluIleLeuLysGlnValLeuLeuAspGlnHisLysAspThrLeuLysAsp 583
Db 349 AGAGTTTCAGAGCAATTTAGAAAAGTATTTTAAAAAGAAAAGATCATGCTTATATATAT 408
QY 584 GlnLysSerCysValLysTyrLeuLysGluLysCysAsnLysTyrSerArgArgGlyAsp 603
Db 409 GAGCAAAACGTGACGAGCGTGTTCAGAAAGAAATGTATATCTTATATAGAGGAGAAAG 468
QY 604 AspArgPheSerPheValCysValPheGlnAsnAlaThrCysGluLeuMetValLysAsp 623
Db 469 AATTCATTGTGTGTTTCATGTGCTTTCACAGAGAAACATGTGAATATATGTGTATCCGT 528
QY 624 ValLysAspArgCysGluValPheLysLysAsnLysIleLysAlaSerTyrIleGluPhe 643

```

Db 529 ACAAAAGATGATGTTTATTATTAAGTCGCAACATGAGATGAAAAATTGTAAGAA 588
 Qy 644 LeuGluAenAnthrAenlylIeThrThreugluAenCyseProserThriIeThr 663
 Db 589 ATTGGAAAAGAAAAAGCAATGAAACAGCACTCGAAGAACTCTGCACACATGGGGCCGA 648
 Qy 664 TyrCyseAnthrPheSerProAnCysePro-----GlyLeu 675
 Db 649 CATGTGCACCAACTTATGAGAAATTGTCCGGATGACTTGAAGAAAAAGAAATGGCAAT 708
 Qy 676 ThrlyeGluAenSerCyseThrlyeIlelylelyshIaArgGluProPheThrlyeArylys 695
 Db 709 GACAAATATCATTAATCTGTAAGAACTCGATGAAAAATGACGTGATACCTTAAAGATGG 768
 Qy 696 AleuGluAenAlaLeuLyValGluLeuGlnlyLyseThrAenlyPheSerlyeCys 725
 Db 769 AAATTAAAGGAGGAGCTGACTCATCTGTTGAAAGCAGCTTAAAGATTAAGATGAATGT 828
 Qy 716 GluProAlaLeuLyAArgTyrCyseThr---ValAlaGlyAenValAenAenAlaSerIle 734
 Db 829 AAAAAACATTAGAAAAAGCCGTTCAGCTGAGTTCGCAAAATATGAAACATTAAAAATCTG 888
 Qy 735 SerGlyLeuCyseLyValAenThrlyeAenPheAnSerGlylyeSerAenGluAenPalAArg 754
 Db 889 CTTAGTATTGTGGAGAAATGACAAAGGAACT----- 921
 Qy 755 LyseGluLeuCyseGluLyValGluValGluGluGlnCyseLyValAleuProThr 774
 Db 922 -----GTTGCGAAAAATTGTAATAAAAGAGATGCTCTTAAAGAAC 975
 Qy 775 GluLeuGlnGlnProAlaAlaAenPheLyseAenPheThrlyeGluGluLeuLylys 794
 Db 976 GGAATGAAATTAAGGAAAGATGAGTTCGACAAAGATGAAAAAGATACGATCGCCTTAA 1035
 Qy 795 LyseArgAlaGluGluAlaMetAenlyseSerleuValleuSerleuIlelylys--- 813
 Db 1036 AAGCGCGCAGAAAGATCTACAAAGAAAGCTTATGCTATCAAGACCTAGCAAACT 1095
 Qy 813 ----- 813
 Db 1096 GTAATGCCAAGTCGCGCAAGATGCAAGTCTTCAAGCAAGTATTACAAACGTAACAACA 1155
 Qy 814 -----AenGluSerAenValSerlyseAenSerlyeAenlyAenlyAenAlaVal 831
 Db 1156 GAATCAGGGTATCATCAGGTCGCATCATCACCAAGGCGCACCATCGACACCAACA 1215
 Qy 832 SerAenGlyLeu-----GlnAenPheThrThr 839
 Db 1216 CAAAAATGAAACGCCAGCCACACAGTGAAGCACACAGGACACCAAGCAGTGAAGACG 1275
 Qy 840 LyseHisValys-----IleLeuArgArgGlyVallyAenPheValSerValThrGluLeu 857
 Db 1276 GGCCTCGCAAAACTTGACTGCTTAAAGACATATGTACTGAAGATATCAGAAAGCA 1335
 Qy 858 GluAlaLyValAenPheAenPheAlaAlaGluValPheGlyArgTyrValAenPheLyseGlu 877
 Db 1336 GAGGTCAAAGATTTGATGCAACAACATGACATTTGAGTTTGAATTTGAAGAA 1395
 Qy 878 ArgCyseAnlylyseGluSerAenPheArgIlelyseGluAenPheCyseLyseAenPheGluGlu 897
 Db 1396 GAATGTAAAGCTTAAAGTATGATTTGCGGTTTAAAGAGATGTAAGAAACTGAACA 1455
 Qy 898 ValCyseLyseLyseIleAenlyAenlyCyseArgAenleuysProleuGluVallyeProHis 917
 Db 1456 GCTTGTAAAGAAATGAAAGATTAATGTA---CTGGAAGCATTAAGATTTGCGCCTCAT 1512
 Qy 918 -----GluThrValThrGluSerThrThrThrThrThrThrThrThrVal--- 934
 Db 1513 CATACAGAGACATTAACAATAAAGTGAAGGAAACACAGACGAAACAAAGACCGTTGAG 1572
 Qy 935 ---AlaAenProLyValA----- 939
 Db 1573 AAGTCGATGACAAAGGCTGATGTAAGACCGTTGAAGAACTGTTACGTAACCAACA 1632

Qy 940 -----ThrGluCyseLyseSerleuGlnThrThrAenPheThrTyrVal 952
 Db 1633 GGAAGTGAAGAAAAAGTAACAGAAAGAGTGTACATGATACAAACACAGATACATGGGTG 1692
 Qy 953 ThrGlnThrSerThrHisThrSerThrThrIleThrSerThrIleThrSerlyeIle 972
 Db 1693 ACAAGCAGCTCATTTGCATACGAGTACGACAAAGTACATCGACGGTGAAGTGAAGTGTG 1752
 Qy 973 ThrleuThrSerThrAArgCyseLyseProThrlyeCyseThrThr-----GlyGluGlu 990
 Db 1753 ACGTTGACCTCGATGCGCAAGTGCAGCCTTACCAATGTAATGTAATGTAATGTAATGTA 1812
 Qy 991 AspAenAlaGly-----AspVallyeProserGlnGlyLeuArgMetSer 1005
 Db 1813 ACAGATTAAGAGAGAGAGAGAGAGAGAGAGATGTAACCAATGACGAGATGAAGAAATAGA 1872
 Qy 1006 GlyTyrAenValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023
 Db 1873 GTTCCTGATATGATTAATAATATGTTGTGGAGTATGTTATGCGAATGATG 1926

Search completed: October 29, 2005, 10:12:11
 Job time : 1430 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: October 29, 2005, 08:57:56 / Search time 7490 Seconds
(without alignments)
5198.898 Million cell updates/sec

Title: US-10-654-416-14
Perfect score: 5411
Sequence: 1 MARAVKQAGTQNSIDEH.....MSGNVNRGVIVAVISFMI 1023

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 1902134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10654416/runat_26102005_101023_21745/app_query.faeta_1.1159
-DB=EST -QFMT=faeta+ -SUPP=+p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -XGAPOP=2000000000
-USER=US10654416 @CGN 1.1 5514 @runat_26102005_101023_21745 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUDRRY -NEG SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hc: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_g881: *
9: gb_g882: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	460.5	8.5	746	2	AM334782
C 2	455	8.4	789	2	AM335265
C 3	453.5	8.4	703	2	AM331960
C 4	450.5	8.3	736	2	AM333066
C 5	443.5	8.2	693	2	AM334544
C 6	441.5	8.2	742	2	AM335497
C 7	430	7.9	737	2	AM334880
C 8	429	7.9	725	2	AM333034
C 9	427	7.9	705	2	AM334533

C 10	415	7.7	719	2	AM334763	AM334763	S37F11	AG
C 11	409.5	7.6	647	2	AM334174	AM334174	S31D8	AGS
C 12	402	7.4	649	2	AM331911	AM331911	S2C12	AGS
C 13	394	7.3	733	2	AM332556	AM332556	S9H7	AGS-
C 14	393	7.3	710	2	AM332531	AM332531	S9E12	AGS
C 15	389	7.3	800	2	AM334916	AM334916	S40G3	AGS
C 16	389	7.2	572	2	AM333173	AM333173	S44A3	AGS
C 17	387.5	7.2	699	2	AM333354	AM333354	S20F5	AGS
C 18	384	7.1	617	2	AM332662	AM332662	S11C9	AGS
C 19	380.5	7.0	593	2	AM333429	AM333429	S21F6	AGS
C 20	376.5	7.0	739	2	AM334513	AM334513	S35G10	AG
C 21	374.5	6.9	475	2	AM334345	AM334345	S33E10	AG
C 22	367	6.8	628	2	AM331939	AM331939	S2F3	AGS-
C 23	364	6.7	739	2	AM334475	AM334475	S35C11	AG
C 24	363	6.7	654	2	AM334208	AM334208	S31G2	AGS
C 25	363	6.7	761	2	AM334463	AM334463	S35B3	AGS
C 26	362.5	6.7	743	2	AM334196	AM334196	S31F5	AGS
C 27	359.5	6.6	694	2	AM334689	AM334689	S36G7	AGS
C 28	358	6.6	714	2	AM334897	AM334897	S40B6	AGS
C 29	350.5	6.5	566	2	AM332050	AM332050	S3E3	AGS-
C 30	343	6.3	716	2	AM334785	AM334785	S37H11	AG
C 31	331	6.1	540	2	AM334225	AM334225	S32A3	AGS
C 32	324.5	6.0	682	2	AM335277	AM335277	S45C1	AGS
C 33	318	5.9	559	2	AM333479	AM333479	S2C3	AGS
C 34	317	5.9	576	2	AM333182	AM333182	S18D8	AGS
C 35	316	5.8	582	2	AM332456	AM332456	S8F12	AGS
C 36	312	5.8	624	2	AM334653	AM334653	S38G9	AGS
C 37	307	5.7	658	2	AM334501	AM334501	S35F7	AGS
C 38	304	5.6	584	2	AM333103	AM333103	S17C12	AG
C 39	304	5.6	672	2	AM333158	AM333158	S4H11	AGS
C 40	300.5	5.6	527	2	AM334227	AM334227	S32A8	AGS
C 41	298.5	5.5	563	2	AM334815	AM334815	S39E9	AGS
C 42	297.5	5.5	743	2	AM335500	AM335500	S47H10	AG
C 43	297	5.5	603	2	AM332931	AM332931	S15A5	AGS
C 44	292	5.4	575	2	AM332985	AM332985	S15G12	AG
C 45	286	5.3	612	2	AM334018	AM334018	S29C10	AG

ALIGNMENTS

RESULT 1
AM334782/C 746 bp mRNA linear EST 31-JAN-2000
LOCUS S37H4 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
DEFINITION AM334782
ACCESSION AM334782
VERSION AM334782.1 GI:6831048
KEYWORDS EST.
SOURCE
Pneumocystis carinii
Pneumocystis carinii
ORGANISM
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE
1 (bases 1 to 746)
Smullan,A.G., Arnold,V., Weise,M., Wunderlich,J., Staben,C.,
Edman,J.C., Kovacs,J. and Cuelhion,M.
Expressed sequence tags from Pneumocystis carinii
JOURNAL
Unpublished (2000)
COMMENT
Contact: Staben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: staben@pop.uky.edu.
Location/Qualifiers
1..746
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_id="AGS-1"
/note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI;
P. carinii organisms (3x10e9) from a single rat (99-1-6,"

FEATURES
source

sacrificed on 3/17/99) at Cincinnati VA facilities. Total RNA extracted using RNeasy spin columns (Qiagen) under conditions described by the vendor, Stratagene. Further details see www.uky.edu/Project/Pneumocystis/

ORIGIN

Alignment Scores:	
Pred. No.:	8,44e-36
Score:	460.50
Percent Similarity:	62.21%
Best Local Similarity:	43.32%
Query Match:	8.51%
GB:	2
Length:	746
Matches:	94
Conservative:	41
Mismatches:	51
Indels:	31
Gaps:	5

US-10-654-416-14 (1-1023) X AW334782 (1-746)

Qy	838	ThrThyLhnhIsvAllye-----lleuwaTgTgVgLyVlyAseVseValThr	855
Db	731	ACAACGATATATCAAAACCTTGACCTGTTAAAGACATATGTAGAGGAGGTATCA	6723
Qy	856	GluleGluValLysAlaPheAsePleuAlaIaGluValPheGlyArTyValAsPleu	8757
Db	671	GAAGCAGAGGTAAAGCATTTGTGTGAGCGAGCAGATAGCATTTGGAGCTGTATTGGAAATTG	61212
Qy	876	LysGluuArgCyAseAnlyVleuGluSeArPcyVatglllelyVgluAsPcyVlyAsePleu	8955
Db	611	AAAGAGAAATGCGAGCCCTTACACATAGATTCGCGTTTAAAGAGATTTGGAGAAATCT	5523
Qy	896	GluGluValCyVelyVlyleAnlyVsalCyVatAseAnleuVlyProleuGluVallye	9151
Db	551	AAAAAGTTGGCGAGAAATACACAGTTATTCGAAGAAATAAACATTGAAATTAG	4923
Qy	916	ProHieGluThrValThrglu-----	9223
Db	491	CCTCATCATATCAGAGACGCAAAAAGAAATCTCAACCATCAGCAGCACAACAGACC	4323
Qy	923	-----SerThrThrThrThrThrThrThrThrThrThrValAlaAsPPro-----	9323
Db	431	ACTACCAACACTATCTACGACTACGACTATACGAGACTACTACACACAAACCAAGCCGGAAAT	3723
Qy	938	-----LysAlaThr---GluCyVlySeSerLeuGlnThrThAspThrTriValThrgln	9543
Db	371	GGAGAAAAGTAAACAGAAAGGTGTCAAATGATACAGACAGACAGATACATGGGTGACAGAT	31212
Qy	955	ThrSeThrHisThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThr	9743
Db	311	ACCTCATTTGATTCGAGTACGATACGATACGACACGATCAAGCGTACGTCGACAGTACGCTTG	2523
Qy	975	ThrSeThrArgArgCyVelyPProThrlVlyCyVThrTr-----	9873
Db	251	ACGTCGATGGCGAAGCGAACCCTTACCAATGATGACTACTGATTCAAACAAAGAAACACAG	1923
Qy	988	---GlyGluGluAsPAlaIaGlyAsPValVlyProSeGluGlyLeuArgMetSeGly	10003
Db	191	AAAGGAGAGAAAGCAGAAAGAAAGATGTAACCAATATGATGGGATGAAATTAAGATT	1323
Qy	1007	TrpAsnValMetArgGlyValIleValAlaMetValIleSerPheMetIle	1023
Db	131	CCGATATATGATTTAAATATATGTTGTGGAGAGATTTGTTATGGGATATATG	81

RESULT 2

LOCUS	AW335265	789 bp	mRNA	linear	EST 31-JAN-2000
DEFINITION	S45B7 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.				

```

ACCESSION  AW335265
VERSION     AW335265.1  GI:6831884
KEYWORDS    EST.

```

RECORDS	ES1.
SOURCE	Pneumocystis carinii
ORGANISM	Pneumocystis carinii

Eukaryota; fungi; Ascomycota; Pneumocystidomycetes, Pneumocystidaceae; Pneumocystis.

REFERENCE
1. Dabbs 1 CO /89/
AUTHORS
Smulian, A.G., Arnold, J., Weise, M., Wunderlich, J., Staben, C.,

TITLE
JOURNAL
COMMENT

Edman, J.C., Kovacs, J. and Cushion, M.
Expressed sequence tags from *Pneumocystis carinii*
Unpublished (2000)
Contact: Staben C

COMMENT

101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: staben@pop.uky.edu
Location/Qualifiers
1. .789

`/lab host="E. coli"`
`/clone_1lb="AGS-1"`
`/note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI`
`P. carlini organisms (3x106) from a single rat (99-1-6,`
`sacrificed on 3/17/99) at Cincinnati VA facilities.`
`Trizol extracted RNA. Oligo dt priming, strand`
`conditions described by vendor, Stratagene. Further`
`details see www.uky.edu/Project/Pneumocystis/`

ORIGIN

Alignment Scores:	
Pred. No.:	3 34e-35
Score:	452.00
Percent Similarity:	59.29%
Best Local Similarity:	41.59%
Query Match:	8.41%
DB:	2
length:	769
Matches:	94
Conservative:	60
Mismatches:	40
Indels:	32
Gaps:	5

US-10-654-416-14 (1-1023) x AW335265 (1-789)

Qy	823	AsnAlaValSerAsnGlyLeuGlnAspThrThrTrpHisValLys-----IleLeuArg	846
Db	788	GAATGAAATCATTCAGGGCCGATCGAGTCAAAACGAATATGCAAACTTGACATTCGGTTAAA	729
Qy	847	ArgGlyValIleAspValSerValThrGluLeuGlnAlaIleValPheAspLeuAla	866
Db	728	AGAGCATATGTAGATGAGCGTGTATCAGAACAGACGATTAAGATTGATGCACGACG	669
Qy	867	GluValPheGlyArgTyrValAspLeuLysGluArgCysAsnIleLeuGlnSerAspCys	886
Db	668	ATAGCATTGAAATGTGATTTGGATTGAAGAAAGAAATGTAAGCTTATGAACTAGATTGC	609
Qy	887	ArgIleLysGluAspCysLeuAspLeuGlnGluValCysIleValIleAsnIleValCys	906
Db	608	GGTTTAAAGAAAGATTGTTCGAGGTGATGATGATGTTGCAAGAAATTGACACTTATGT	549
Qy	907	ArgAsnLeuLysProLeuGluValLysProHisGluThrValThrGlu-----Ser	923
Db	548	AAA---CTGGAACCATTTGAATTTAAGCTCATCTACACAGACGCAAAAAGAAATCTCA	492
Qy	924	ThrThrThrThrThrThrThrThrThrValAlaAspProLysAlaThr-----	940
Db	491	ACCACATACGACGACCATACGACGACCTACTACACAACTACTACACAGACTACGACGACA	432
Qy	941	-----GluCysLysSerLeu	945
Db	431	ACAACTACTACTACCAAGACGAGGAGTGGAGAAAAGTAACAGAAAGGTGACATGATGA	372
Qy	946	GlnThrThrAspThrTyrValThrGlnThrSerThrIleThrSerThrIleThr	965
Db	371	CACACACAGAAATACATGGGTGACGAGCACGTCATTCGATACGAGTACGACAAACGAGTACG	312
Qy	966	SerThrIleThrSerLysIleThrLeuHisSerThrArgArgCysLysProThrLysCys	985
Db	311	TCGACAGTACGCTCGACAGTACGCTTGACGTCGATGAGAAAGTGCACCAACCTTACCAAAATGT	252
Qy	986	ThrThr-----GlyGlnGluAspAspAlaGlyAspValLys	997

Db 251 ACCACCGATTCAAAACAAGACGATTAAGAGAGAGAGAGAGAGATGTAA 192
Qy 998 ProSerGluGlyLeuArgMetSerGlyTTPanValMetArgGlyValIleValAlaMet 1017
Db 191 CCAAAATGCGGATGAAATAAGAGTTCTGATATGATTAATAATGTTGGAGTG 132
Qy 1018 ValIleSerPheMetIle 1023
Db 131 ATTGTATGCGGATGATG 114

RESULT 3
LOCUS AM331960 703 bp mRNA linear EST 31-JAN-2000
DEFINITION S1B8 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
ACCESSION AM331960
VERSION AM331960.1 GI:6828226
KEYWORDS EST.
SOURCE Pneumocystis carinii
ORGANISM Pneumocystis carinii
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 703)
Smullan,A.G., Arnold,J., Weise,M., Wunderlich,J., Straben,C.,
Edman,J.C., Kovacs,J. and Cushion,M.
Expressed sequence tags from Pneumocystis carinii
JOURNAL Unpublished (2000)
COMMENT Contact: Straben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: straben@pop.uky.edu.
Location/Qualifiers
1. 703
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_1b="AGS-1"
/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
P. carinii organisms (3x10e9) from a single rat (99-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Trizol extracted RNA. Oligo dt priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN
Alignment Scores:
Pred. No.: 3.95e-35 Length: 703
Score: 453.50 Matches: 92
Percent Similarity: 63.00% Conservative: 34
Best Local Similarity: 46.00% Mismatches: 41
Query Match: 8.38% Indels: 33
DB: 2 Gaps: 5

US-10-654-416-14 (1-1023) x AM331960 (1-703)

Qy 854 ValThrGluLeuGluValAlaValPheAspLeuValAlaGluValPheGlyArgTyrVal 873
Db 688 GTATCAAGAGGAGAAAGTAAAGCATTTGATGCAACGACGATGACATGCTTATTTG 629
Qy 874 AspLeuValGluValArgValAsnValLeuGluSerAspCysArgIleValAspCysVal 893
Db 628 GAATTAAGAGGAAATGCAATGCTTTAGAACTAGATTGCGGTTTAAAGAGATTGTTCA 569
Qy 894 AspLeuGluValValCysValValIleAsnValAlaCysValGluValPheGluVal 913
Db 568 GATCTT---CAAGTTTCGCGGAGAAATAGCAAAATTATGAAAGAAACCACTTTGAA 512
Qy 914 ValValPheHisGluValThrValThrGlu-----SerThrThrThrThrThrThr 930

Db 511 ATTAAGCTTCATCATATACAGAGACGCAAAAGAAATCTCAACCACTACGACGACACTACG 452
Qy 931 ThrThrThrValAlaAspProlySalThr----- 940
Db 451 ACCAGCACTACCAACAACCTACTACTACGACTACTACGACTACTACTACGACCAACGAC 352
Qy 941 -----GluCysIleSerLeuGlnThrThrAspThrTTPVal 952
Db 391 GGAAGTCAGAGAAACATAACAGAGAGTGTACAAATGATACAGACGACAGATGATGCGGT 332
Qy 953 ThrGlnThrSerThrHisThrSerThrSerThrIleThrSerThrIleThrSerIle 972
Db 331 ACAAGTACGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 272
Qy 973 ThrLeuThrSerThrArgArgCysValProThrValCysThrThr----- 987
Db 271 ACGTTACGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 212
Qy 988 -----GluGluGluAspAspAlaGluValAspValProSerGluGlyLeuArg 1003
Db 211 ACAGATTAAG 158
Qy 1004 MetSerGlyTTPanValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023
Db 157 ATTAAGATTCCTGATATGATTAATAATGTTGAGAGTGTATGATGCGGATGATG 98

RESULT 4
LOCUS AM333066 736 bp mRNA linear EST 31-JAN-2000
DEFINITION S16G12 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
ACCESSION AM333066
VERSION AM333066.1 GI:6829423
KEYWORDS EST.
SOURCE Pneumocystis carinii
ORGANISM Pneumocystis carinii
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 736)
Smullan,A.G., Arnold,J., Weise,M., Wunderlich,J., Straben,C.,
Edman,J.C., Kovacs,J. and Cushion,M.
Expressed sequence tags from Pneumocystis carinii
JOURNAL Unpublished (2000)
COMMENT Contact: Straben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: straben@pop.uky.edu.
Location/Qualifiers
1. 736
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_1b="AGS-1"
/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
P. carinii organisms (3x10e9) from a single rat (99-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Trizol extracted RNA. Oligo dt priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN
Alignment Scores:
Pred. No.: 8.57e-35 Length: 736
Score: 450.50 Matches: 88
Percent Similarity: 63.16% Conservative: 44
Best Local Similarity: 42.11% Mismatches: 48
Query Match: 8.33% Indels: 29
DB: 2 Gaps: 4

```
US-10-654-416-14 (1-1023) x AM333066 (1-736)

Oy 844 ILeuArGrgLYValIyAspValSerValThrgIuLeuGluAlaIyAlaPheAsp 863
    ::::::::::::::::::::|
Db 725 CTCGTAAAGAGAGCATATGATGAGAGAGTATCAAGACAGAGCTAAAGCATTTGAT 666
Oy 864 LeuAlaIaGluValPheGlyArGryValAspleuLySGIuArGyCyAsnLyLeuGlu 883
    ::::::::::::::::::::|
Db 665 GCACGCGCCGTAGCATGTGAGATGTATTGGATTGAAAGAGAGATCGACGCTTTACAA 606
Oy 884 SerAspCyAsnGlyIleLyGluAspCyAspLeuGluValCylValylLeuAsn 903
    ::::::::::::::::::::|
Db 605 CTAGATTTCGGTTTAAAGAGAGATGTCAAGATCTAAACAGCTTGCGAAAAATTAGAC 546
Oy 904 LySAlaCyAsnLyAsnLeuLyPProLeuGluValIyLyPProHISGluThrValThrgIu--- 922
    ::::::::::::::::::::|
Db 545 GAGTTATCCGAGAGAAATAAACATTAAAGAAATTAAGCTCATCAACAGAGACGAAAAA 486
Oy 923 -----SerThrThrThrThr 927
    ::::::::::::::::::::|
Db 485 GAAATCTCAACCACTACGACGACCACTACGACACTACCAACCACTACCACTACCA 426
Oy 928 ThrThrThrThrThrThrValAlaAspPro-----LyAlaLeThr---GluGys 942
    ::::::::::::::::::::|
Db 425 ACGAGCAACACTACTACTACCAACCAAGCCGGAGAGTGAGAGAAAGTAACAGAAAGTGT 366
Oy 943 LySerLeuGlnThrThrAspThrTrpValThrgIuThrSerThrHisIleThrSerThrSer 962
    ::::::::::::::::::::|
Db 365 ACAATGATACAAACAACAGATACATGGGTAAAGCGCTACGTCTCATCAACAGATCGACA 306
Oy 963 ThrIleThrSerThrIleThrSerLySileThrLeuThrSerThrArGArCyAspLeuPro 982
    ::::::::::::::::::::|
Db 305 ACGAGTACGTGACAGTACGTGACAGTACGTGACGTGACGTGACGTGACGTGACGTGAC 246
Oy 983 ThrLySArGrgLYValIyAspValSerValThrgIuLeuGluAspAlaGly 994
    ::::::::::::::::::::|
Db 245 ACCAATATACCAACCGATTCAAAACAAGACAGATTAAGAGAGAGAAAGAGAGAGAA 186
Oy 995 AspValLyPProSerGluGlyLeuArGmeSerGlyTrpAsnValMeArGlyValIle 1014
    ::::::::::::::::::::|
Db 185 GATGTAAACCGAATGATGGATGAAATTAAGAGTCTGATATGATTAATAATATGTTG 126
Oy 1015 ValAlaMeValIleSerPheMetIle 1023
    ::::::::::::::::::::|
Db 125 TTGGGAGTGTATTGCGGATGATG 99

RESULT 5
AM334544/c 693 bp mRNA linear EST 31-JAN-2000
LOCUS S35B1 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
DEFINITION AM334544
VERSION AM334544.1 GI:6830901
KEYWORDS EST.
SOURCE Pneumocystis carinii
ORGANISM Pneumocystis carinii
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 693)
AUTHORS Smilian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
Edman,J.C., Kovacs,J. and Cushion,M.
Expressed sequence tags from Pneumocystis carinii
JOURNAL Unpublished (2000)
COMMENT School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 1161
Fax: 606 257 1717
Email: staben@pop.uky.edu.
location/Qualifiers
FEATURES
Source 1..693
```

```
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_lib="AGS-1"
/vector="Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
P. carinii organisms (3x10e9) from a single rat (59-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Tizol extracted RNA. Oligo dt priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN

Alignment Scores:
Pred. No.: 4,01e-34 Length: 693
Score: 443.50 Matches: 91
Percent Similarity: 61.24% Conservative: 37
Best Local Similarity: 43.54% Mismatches: 48
Query Match: 8.20% Indels: 33
DB: 2 Gaps: 5

US-10-654-416-14 (1-1023) x AM334544 (1-693)

Oy 845 LeuArGrgLYValIyAspValSerValThrgIuLeuGluAlaIyAlaPheAspLeu 864
    ::::::::::::::::::::|
Db 692 GTTAAAGGCGATATGTAGCTGACGAGTATCAAGACAGAGTAAAGCATTTGATGCA 633
Oy 865 AlaAlaGluValPheGlyArGryValAspleuLySGIuArGyCyAsnLyLeuGluSer 884
    ::::::::::::::::::::|
Db 632 ACGAGATAGATTTGAGAGCTGATTTGGAATTTGAAAGAGAAATGCAATGCTTTAGACTA 573
Oy 885 AspCyAsnGlyIleLyGluAspCyAspLeuGluGluValCylValylLeuAsnLy 904
    ::::::::::::::::::::|
Db 572 GATTCGGCTTTTAAAGAGATTTGTTCAAGATCTT---CAAGTTTCGGAGAAATGTCAAG 516
Oy 905 AlaCyAsnLySileLyGluAspProLeuGluValIyPProHISGluThrValThrgIu----- 922
    ::::::::::::::::::::|
Db 515 TTATGCAAGAAATTAAGACCATTTGAAATTAAGCTCTCATCAACAGAGACCAAAAAAGAA 456
Oy 923 -----SerThrThrThrThrThr 930
    ::::::::::::::::::::|
Db 455 ATCTCAACCACTACGCGGAGACCACTACGACGACCACTACCACTACCTACGATGCA 396
Oy 931 ThrThrThrValAlaAspProLyS-----AlaThrGluCyLyS 943
    ::::::::::::::::::::|
Db 395 ACGACTACTACTACGACCAACCAAGCAGAGAGTGGAGAAAGTAACAGAAAGTGTACA 336
Oy 944 SerLeuGlnThrThrAspThrTrpValThrgIuThrSerThrHisIleThrSerThrSerThr 963
    ::::::::::::::::::::|
Db 335 ATGATACAGACAGATACATGGGTGCAAGTACATTTGATACATGACAGTACGACAACG 276
Oy 964 ILeThrSerThrIleThrSerLySileThrLeuThrSerThrArGArCyAspLeuProThr 983
    ::::::::::::::::::::|
Db 275 AGTACGTGACAGTACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 216
Oy 984 LyCySArGrgLYValIyAspValSerValThrgIuLeuGluAspAlaGly 994
    ::::::::::::::::::::|
Db 215 AAAAGTACCAACCGATTCAAAACAAGACAGATTAAGAGAGAGAAAGAAAGAA----- 162
Oy 995 AspValLyPProSerGluGlyLeuArGmeSerGlyTrpAsnValMeArGlyValIle 1014
    ::::::::::::::::::::|
Db 161 GAAATTAACCAATATGATGGATGAAATTAAGAGTCTGATATGATTAATAATATGTTG 102
Oy 1015 ValAlaMeValIleSerPheMetIle 1023
    ::::::::::::::::::::|
Db 101 TTGGGAGTGTATTGCGGATGATG 75

RESULT 6
AM335497/c 742 bp mRNA linear EST 31-JAN-2000
LOCUS S47H6 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
DEFINITION AM335497
VERSION AM335497.1 GI:6832118
```

KEYWORDS EST.
SOURCE Pneumocystis carinii
ORGANISM Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 742)
AUTHORS Smullen,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
Edman,J.C., Kovacs,J. and Cushion,M.
TITLE Expressed sequence tags from Pneumocystis carinii
JOURNAL Unpublished (2000)
COMMENT Contact: Staben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: staben@pop.uky.edu.
Location/Qualifiers
1. 742
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_lib="AGS-1"
/note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI;
P. carinii organisms (3x10e9) from a single rat (99-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Trizol extracted RNA. Oligo dt priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN

Alignment Scores:
Pred. No.: 7,13e-34 Length: 742
Score: 441.50 Matches: 92
Percent Similarity: 60.68% Conservative: 33
Best Local Similarity: 44.66% Mismatches: 50
Query Match: 8.16% Indels: 31
Gaps: 5

US-10-654-416-14 (1-1023) x AW335497 (1-742)

QY 847 ArgGlyValIysAspValSerValThrGluLeuGluAlaValAspLeuAla 866
DB 741 AGACCATATGATGAGAGATGACAGACGATTAAGCATTTGACGACAGC 682
QY 867 GluValPheGlyArgValAspLeuLysGluArgCysAsnLysLeuGluSerAspCys 886
DB 681 GTACCATTTGGAATGTATTTGGAATTTGAAAGAGAAATGTAAGCTTTAGATTGCTC 622
QY 887 ArgGlyLeuGluAspCysValAspLeuGluValCysLysLysLeuAsnLysAlaCys 906
DB 621 GGTTTTAAAGAGATTTGTTGATCTT---CAAGTTTGTGAGAAATAGACACGTTATGC 565
QY 907 ArgAsnLeuLysProLeuGluValLysProHisGluThrValThrGlu-----Ser 923
DB 564 GGAGCAATTAACCAATTAAGAAATTAAGCTCATATACAGACGCAAAAAGAAATCTCA 505
QY 924 ThrThrThrThrThrThrThrThrThrThrValAlaAspProLysAlaThr----- 940
DB 504 ACCCACTACGACGACCACTACGACGACACTACCACTACTACGACGACGACGACA 445
QY 941 -----GluCysLysSerLeu 945
DB 444 ACAACTACTACTACCAAGCAGGAGAGTGGAGAAAAGTACAGAAAGAGTACATGATA 385
QY 946 GlnThrThrAspThrTrpValThrGlnThrSerThrHisThrSerThrSerThrIleThr 965
DB 384 CAAACCAACAGATACATGGGTAAACGTAAGTCATTTGATACAGATACACAAACAGTACG 325
QY 966 SerThrIleThrSerValIleThrLeuThrSerThrArgAGCysLysProThrLysCys 985

DB 324 TCAACGGTGAAGTCGACAGTGAACGTTGACGTCGATGAGAAAGTGCAAACCTACCAATGT 265
QY 986 ThrThr-----GlyGluGluAspAspAlaGlyAspValLys 997
DB 264 ACCACCGATTTCAACGACAGACAGATTAAGAGAGAGATGAAGAA---GAAGTAAA 208
QY 998 ProSerGluGlyLeuArgMetSerGlyTrpAsnValMetArgGlyValIleValMet 1017
DB 207 CCAATATGATGAGTGAATAAATTAAGAGTTCCTGATATGATTAATAATATGTTGGAGAGTG 148
QY 1018 ValIleSerPheMetIle 1023
DB 147 ATTGTTATGGAAATGATG 130

RESULT 7
AW334880/c 737 bp mRNA linear EST 31-JAN-2000
LOCUS AW334880
DEFINITION S40C12 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
ACCESSION AW334880
VERSION AW334880.1 GI:6831237
KEYWORDS EST.
SOURCE Pneumocystis carinii
ORGANISM Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 737)
AUTHORS Smullen,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
Edman,J.C., Kovacs,J. and Cushion,M.
TITLE Expressed sequence tags from Pneumocystis carinii
JOURNAL Unpublished (2000)
COMMENT Contact: Staben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: staben@pop.uky.edu.
Location/Qualifiers
1. 737
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_lib="AGS-1"
/note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI;
P. carinii organisms (3x10e9) from a single rat (99-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Trizol extracted RNA. Oligo dt priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN

Alignment Scores:
Pred. No.: 1.04e-32 Length: 737
Score: 430.00 Matches: 93
Percent Similarity: 56.35% Conservative: 49
Best Local Similarity: 36.90% Mismatches: 96
Query Match: 7.95% Indels: 14
Gaps: 5

US-10-654-416-14 (1-1023) x AW334880 (1-737)

QY 553 CysHisThrLeuGluGluGlnCysAsnArgLeuGlyThrThrGluIleLeuLysGlnVal 572
DB 737 TGTATTAATTTGAATGATGACCTGCAATATTTAGAACTTCAGAAAGAACTCAACAGACATA 678
QY 573 LeuLeuAspGluHisLysAspThrThrLeuLysAspGlnGluSerCysValLysTrpLeuLys 592
DB 677 TTTTAAAGAGAGAAAGAAATATTTATTAAGAAATATTTGAAAAGCCTTAAA 618
QY 593 GluLysCysAsnLysTrpSerArgArgGlyAspAspArgPheSerPheValPhe 612

Db	61.7	GAATAATGTGATGGGATATTATTTAAAAAGAGACGAATCTATTAGATTCAATGCTTTTA	558
Qy	61.3	GLHsenaIaThrCySGluLeuMetValIySaSpValIySaSpArgCySGluValaPheIyS	632
Db	55.7	CAGAAAGAAATATGTGCAATATAGTCTCAAGACAGAAACAAACAGTGTCAATCTTTAA	498
Qy	63.3	LYSaenIIeIySaIaSerTyrIleTleGIuPheLeuGIuSaenThrSaenIySaIeThr	652
Db	49.7	GAGAACATATGAAAAACAGAAAAATTTGGAT---AAAGAAACCGAAAAAATGAAACA	441
Qy	65.3	ThIeugIuAArgAsnCySProSerTPrhIaThrTyrCySaenArgPheSerProAsnCyS	672
Db	44.0	TCACTTAAAGAAATATGTATTATATGAGAGTCATATTCGAATCAGCTTATGGAATTTG	381
Qy	67.3	ProGIyLeuThrIySGluSaenSer-----CyThrIySaIeIySaIyS	686
Db	38.0	CCAAAGCTATATGAAAGAAAGAAAGTAAATGACACTGATGAGTGTGTTGAAACTTAAGA	321
Qy	68.7	HisArgIuPProPheTyrIyAArgIySaIaLeuGIuSaenPalaleuIyValaGIuLeuGI	706
Db	32.0	AAATGTACCTATATCTTGGAAGAAAGACCTTGATGATGAATTAATCACTCAATATTGAA	261
Qy	70.7	GIyLyLeuThrIyAspIyLeuSerIySCySGluPProAlaLeuIySaArgTyrCyThrValaI	726
Db	26.0	GGAATATTAAAGTAATATGCTTAATAGTAAACTAGGCTTAAAGAGCTATTGCACTGAAG	201
Qy	72.7	GIySaenValaSaenIIaSerIleSerGIyLeuCyIySaIaSaenThrIySaSpSaenSer	746
Db	20.0	AAAACGTGACGAAAGAAATATATCTTTAAAAATCAGTGT-----AATATCGATGAAT	150
Qy	74.7	GIyIySaSerIyAspGIuSaenPalaleuArgIySGluLeuGIySaIyLeuValIySGluVal	766
Db	14.9	--AAGACCGTTCAAGAT-----GAAATTTGTGAAGATTAGTCAAAAAAATATACG	102
Qy	76.7	GIuGIuGIySaIySaIaLeuProThrGIuLeuGIuGIuPProAlaSaenIySaSp	786
Db	10.1	AAAGATCTCCAACCTTAATAATTAATTGAAAAAGCGCGAAAAAGATTTACAACAA	42
Qy	78.7	TyrIySaThrTyrGIuGIuLeuIySaIySaArgAlaGIu	798
Db	4.1	GAATAAGATATTAATAAAAAAAAAAAAAAACTCGAG	6
RESULT 8			
AW333034/c			
LOCUS	AW333034	725 bp	mRNA linear EST 31-JAN-2000
DEFINITION	SLSD10 AGS-1	Pneumocystis carinii	cdNA 3', mRNA sequence.
ACCESSION	AW333034		
VERSION	AW333034.1	GI:6829391	
KEYWORDS	EST.		
SOURCE	Pneumocystis carinii		
ORGANISM	Pneumocystis carinii		
REFERENCE	Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;		
AUTHORS	Pneumocystidaceae; Pneumocystis.		
TITLE	1 (bases 1 to 725)		
JOURNAL	Smulian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,		
COMMENT	Edman,J.C., Kovacs,J. and Cushion,M.		
	Expressed sequence tags from Pneumocystis carinii		
	Unpublished (2000)		
	Contact: Staben C		
	School of Biological Sciences		
	University of Kentucky		
	101 Morgan Building, University of Kentucky, Lexington, KY		
	40506-0225, USA		
	Tel: 606 257 2161		
	Fax: 606 257 1717		
	Email: staben@pop.uky.edu.		
FEATURES	Location/Qualifiers		
source	1..725		
	/organism="Pneumocystis carinii"		
	/mol_type="mRNA"		
	/db_xref="taxon:4754"		
	/lab_host="E. coli"		
	/clone_lib="AGS-1"		

ORIGIN		Alignment Scores:	
Pred. No.:	1.28e-32	Length:	725
Score:	429.00	Matches:	88
Percent Similarity:	61.62%	Conservative:	34
Best local Similarity:	44.44%	Mismatches:	48
Query Match:	7.93%	Indels:	28
DB:	2	Gaps:	5
US-10-654-416-14 (1-1023) x AM333034 (1-725)			
QY	854 ValMhrgIuLeuGluValAlaLysAlaPheAspLeuAlaIaGluValPheGlyArgTyrVal	873	
DB	709 GTATCAGAGAGAGAGGTAAAGGATTTGATGCAACGACGGTAGCATTTGAAATGTAATTTG	650	
QY	874 AspLeuLysGluArgCysAsnLysLeuGluIserAspCysArgIleLysGluAspCysLys	893	
DB	649 GAATTGAAGAAGGAATCCATTCCTTTGAACCTAGATTCGCGTTTAAAGAGATGGAG	590	
QY	894 AspLeuGluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGlu	913	
DB	589 GAATCTAAACACGCTTTAAAGAAATAGAAAGAGCTTATCAAGAAATATGATCATTTAAA	530	
QY	914 ValLysProHISGlu	918	
DB	529 GTTACGGCTCATCTATACAGAGAAATATACAGAAATCTCAACCACTACGACGACCACTACT	470	
QY	919 ---ThrValThrGluSerThrThrThrThrThrThrThrThrThrValAlaAspPro	937	
DB	469 ACGACCACTACTACGACTACTACACAGACTATCTACGACCACTACTACTACACCAAGCG	410	
QY	938 -----LysAlaThr---GluCysLysSerLeuGlnThrAspThrTyrVal	952	
DB	409 GGAAGTGAGAGAAAGTAAACAGAAAGAGTATGATCAATGATACAGACAGACAGATACATGGGTG	350	
QY	953 ThrGlnThrSerThrHISThrSerThrSerThrIleThrSerThrIleThrSerLysIle	972	
DB	349 ACGAGCAGCTCATTTGATACGAGAGTACACACACGATCGTCACGGTGAACGTCACAGCTG	290	
QY	973 ThrLeuThrSerThrArgArgCysLysProThrLysCysThrThr-----GlyGluGlu	990	
DB	289 ACGTTGACTTGCATGCCGAAAGTCGAAAGCTTACCAAAATGTACTACTGATTCAAGCAGAGAG	230	
QY	991 AspAspAlaGly-----AspValLysProSerGluGlyLeuArgMetSer	1005	
DB	229 ACAGATTAAGAGAGAGAGAGAGAGAAAGATGTAATAACGAAATGATGGGGGTGAATAATAGA	170	
QY	1006 GlyTyrAsnValMetArgGlyValIleValAlaMetValIleSerPheMetIle	1023	
DB	169 GTTCTCATATGATTAAATAATATGTTGTTGGAGCTGATTGTATGGAGATGATG	116	
RESULT 9			
LOCUS	AM334533	705 bp	mRNA linear EST 31-JAN-2000
DEFINITION	S36A7 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.		
ACCESSION	AM334533		
VERSION	AM334533.1	GI:6830890	
KEYWORDS	EST.		
SOURCE	Pneumocystis carinii		
ORGANISM	Pneumocystis carinii		
REFERENCE	Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;		
AUTHORS	Pneumocystidaceae; Pneumocystis.		
	1 (bases 1 to 705)		
	Smilian,A.G., Arnold,J., Weise,M., Wunderlich,J., Steben,C.,		
	Edman,J.C., Kovacs,J. and Cushman,M.		
	Expressed sequence tags from Pneumocystis carinii		
TITLE			

JOURNAL
COMMENT

Unpublished (2000)
Contact: Staben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY
40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: staben@pop.uky.edu

FEATURES
SOURCE

Location/Qualifiers
1..705
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_lib="AGS-1"
/note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI;
P. carinii organisms (3x10e9) from a single rat (99-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Tritzol extracted RNA. Oligo dt priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/Project/Pneumocystis/"

ALIGNMENT SCORES:

Pred. No.: 1,956-32 Length: 705
Score: 427.00 Matches: 86
Percent Similarity: 61.03% Conservative: 33
Best Local Similarity: 44.10% Mismatches: 46
Query Match: 7.89% Indels: 30
DB: 2 Gaps: 4

US-10-654-416-14 (1-1023) x AM334533 (1-705)

858 GluAlaAlaValaPheAspLeuAlaGluValaPheGlyArgTyrValaAspLeuLysGlu 877
702 GAGGTAAAGCATTTGATGCAACGACGATGATGATGATGATGATGATGATGATGATGAT 643
878 ArgCysAlaLeuLysGluSerAspCysArgLysGluAspCysArgLysGlu 897
642 GAATGTAAAGCTTTAGAACTGATGCGTTTAAAGATGTTGAGTGTGATGAT 583
898 ValCysAlaLeuLysGluSerAspCysArgLysGluAspCysArgLysGlu 917
582 GTTTCGAAAGAAATGACACTTATGTAA--CTGACACATTTGAATTAACCTCAT 526
918 GluThrValThrGlu-----SerThrThrThrThrThrThrThrThrThrVal 934
525 CATACAGAGAGCAAAAGAAATCTCAACCACTACGACGACCACTACGACGAC 466
935 AlaAspProLysAlaThr----- 940
465 ACAACTACTACGACGACGACGACCAACTACTACTACGACGACGACGACGACGAC 406
941 -----GluCysLysSerLeuGlnThrThrAspThrTrpValThrGlnThrSer 956
405 AAAGTAACAGAGAGTGTACATATACACGACGACGACGACGACGACGACGACGAC 346
957 ThrHisThrSerThrSerThrThrThrThrThrThrThrThrThrThrThrSer 976
345 TTGCATACGATACGACGACGACGACGACGACGACGACGACGACGACGACGAC 286
977 ThrTrpArgCysLysProThrThrLysCysThr-----ThrGly 988
285 ATGAGAAAGTGCACGCTTACCAATGTACCCCGATTCAACCAAGACAGATTAAGGA 226
989 GluGluAspAlaPheValaAspValaLysProSerGluGluLysArgMetSerGlyTrpAsn 1008
225 GAAAGAGAGAGAGAGAGATGTAAACCAATGATGAGATGAAATTAAGAGTTCCGAT 166
1009 ValMetArgGlyValaLysAlaMetValaLysSerPheMetIle 1023
165 ATGATTAATAATGTTGTTGGAGTGTATGTTATGGGATGATG 121

RESULT 10
AM334763/c

LOCUS AM334763 719 bp mRNA linear EST 31-JAN-2000

DEFINITION S37F11 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.

ACCESSION AM334763

VERSION AM334763.1 GI:6831029

KEYWORDS EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

SOURCE

1..719
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_lib="AGS-1"
/note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI;
P. carinii organisms (3x10e9) from a single rat (99-1-6,
sacrificed on 3/17/99) at Cincinnati VA facilities.
Tritzol extracted RNA. Oligo dt priming, standard
conditions described by vendor, Stratagene. Further
details see www.uky.edu/Project/Pneumocystis/"

ORIGIN

ALIGNMENT SCORES:

Pred. No.: 3,346-31 Length: 719
Score: 415.00 Matches: 95
Percent Similarity: 53.76% Conservative: 48
Best Local Similarity: 35.71% Mismatches: 88
Query Match: 7.67% Indels: 35
DB: 2 Gaps: 6

US-10-654-416-14 (1-1023) x AM334763 (1-719)

60 LysLeuLysGlyPheCysGluAsp---GlyLysAlaAspThrLysCysLysGluLeuLys 78
712 AGAATTAAAGAGTTTGTAAAGATACAGAAAACTTAAGAAATAATGCTGATCTAAAG 653
79 AlaAsnIleGluLysLysCysThrThrLysGlyLysLeuLysGluAlaIleLysLys 98
652 ACAGAGTTTAAGAAAGAGTTTGGACGAAATGAAAGTCCATTTATGATCATTTTCAAT 593
99 LysIleGlnIleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeuPheLeu 118
592 CCT-----TCAGATGAAGAAATGCGCAGAAATATGACCAAAATGTTTATTTTAA 545
119 GluGlyValCysSerLysGluLeuLysAspAspCysAsnThrLeuAsnLysCysTyr 138
544 AGC----- 542
139 GlnLysLysArgAspLysValaAlaGluGluValaLeuLeuArgAlaLeuArgSerAspLe 158
541 CGAAGCAGACGCTCAAAATCTGCGAGATGAAATTTCTTTTAAAGCTTTTAAAGAACAGT 482
158 uAsnGlySerValIleCysGluLysLysLeuLysGluIleCysProValaMetGlyArgG1 178
481 TAAATGACACGAGTAAATGTAAAGAAAGATGAAAGAGTTTGTCCAAATATTAAATTAAGAG 422

Qy	178	uSeAaSpdluLeuThraenLeuCySleuAaNglnlyeGluThrCySlyAaenlleu1	198
Db	421	GAGGACGACGATTAATGTTTCATGATTAATTTTAAACTACGTGTGAAAATCTCGAAA	362
Qy	198	eGluLyuAaPryblybCySgLyThrLeuLyThraPvalSerAlaAlaLeuGlySerPh	218
Db	361	AAAAAAGATGATTTTTTCGAAACCTTTTAAAGCAAAATTAAGTAGAG-----AATGA	311
Qy	218	eLybLybGluThrCySleuGluLeuLeuGlnGlnCnocybryrPhyrrllieGlyAaenCybG1	238
Db	310	ATTAGAGGAAAGATGTCATGAAAGGTG3AAAAAGTTATTTTAAATCAAGTGGCC	251
Qy	238	yAaPaAaPryllellelybCySllleGluLeuGlyLybCySgInGluGlnaenllea1	258
Db	250	TAATTCAGAG-----TGTGATTAATTAATTAAGCAATCGAGGAAAAAATATTAC	200
Qy	258	aTyTwePProPGlyProAaPPhaAaPProThraPProGluAlaThrlleAlaGluAa	278
Db	199	ATACAGAGCTCCAGGATCAGATTTTCTCCAGTTGACCAAAACCAACATTTGTAGAGGA	140
Qy	278	pIlleGlyLeuGlnGluPhyrrlybLybLybValGluLubapGlyValaPheIlleGlyLybAa	298
Db	139	GATTGGCTTGGAAGAAATTTTAAAAAGAGAGAAATTAAGGAATTAATTTGCGGAATCC	80
Qy	298	nHleuLybAaPrybAlaThraAlaLeuAlaLeuAlaLeuIlleGlnAaSerSeryLeuLybLy	318
Db	79	AGGAGAAAGAAAGAGTTGGAGATATCTATTATTATG-----AGTAAAAAAA	32
Qy	318	gLybAaPrybLybGlu	323
Db	31	AAAAAAAAAACTCGAG	16
RESULT 11			
AM334174/c			
LOCUS	AM334174	647 bp	mRNA linear EST 31-JAN-2000
DEFINITION	S31D8 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.		
ACCESSION	AM334174		
VERSION	AM334174.1	GI:6830531	
KEYWORDS	EST.		
SOURCE	Pneumocystis carinii		
ORGANISM	Pneumocystis carinii		
	Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;		
	Pneumocystidaceae; Pneumocystis.		
REFERENCE	1 (bases 1 to 647)		
AUTHORS	Smullen,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,		
TITLE	Edman,J.C., Kovacs,J. and Cushion,M.		
JOURNAL	Expressed sequence tags from Pneumocystis carinii		
COMMENT	Unpublished (2000)		
	Contact: Staben C		
	School of Biological Sciences		
	University of Kentucky		
	101 Morgan Building, University of Kentucky, Lexington, KY		
	40506-0225, USA		
	Tel: 606 257 2161		
	Fax: 606 257 1717		
	Email: staben@pop.uky.edu.		
FEATURES	Location/Qualifiers		
source	1..647		
	/organism="Pneumocystis carinii"		
	/mol_type="mRNA"		
	/db_xref="taxon:4754"		
	/lab_host="E. coli"		
	/clone_lib="AGS-1"		
	/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI		
	P. carinii organisms (3x10e6) from a single rat (59-1-6,		
	sacrificed on 3/17/99) at Cincinnati VA facilities.		
	Triol extracted RNA. Oligo dT priming, standard		
	conditions described by vendor, Stratagene. Further		
	details see www.uky.edu/Project/Pneumocystis/g."		
ORIGIN			
Alignment Scores:			

[illegible]

FEATURES

Location/Qualifiers

1. 649
/organism="Pneumocystis carinii"
/mol_type="mRNA"
/db_xref="taxon:4754"
/lab_host="E. coli"
/clone_id="AGS-1"
/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI; P. carinii organisms (3x10e9) from a single rat (59-1-6, sacrificed on 3/17/99) at Cincinnati VA facilities. Trizol extracted RNA. Oligo dT priming, standard conditions described by vendor, Stratagene. Further details see www.uky.edu/Project/Pneumocystis/"

ORIGIN

Alignment Scores:

Pred. No.:	5.95e-30	Length:	649
Score:	402.00	Matches:	80
Percent Similarity:	62.92%	Conservative:	32
Best Local Similarity:	44.94%	Mismatches:	36
Query Match:	7.43%	Indels:	30
DB:	2	Gaps:	4

US-10-654-416-14 (1-1023) x AW331911 (1-649)

QY 875 LeuLysGluArgCysAsnLysLeuGluSerAspCysArgIleLysGluAspCysLysAsp 894

DB 648 TTGAAAGAGAGAAAGTAAAGCTTTGAACTAGATGGCTTTTACAAAGAGATGTTCCAGT 589

QY 895 LeuGluGluValCysLysLysLeuAsnLysLeuGluSerAspCysArgIleLysGluAspCysLys 914

DB 588 GTTATGATGATGTTTCCAAAGAAATGACACATATGTTAA---CTGGAACCTTTGAAAT 532

QY 915 LysProHisGluThrValThrGlu-----SerThrThrThrThrThrThrThrThrThr 931

DB 531 AAGCTCATCATACAGAGACGCAAAAGAAATCTCAACACTACAGACGACATACAGACG 472

QY 932 ThrThrValAlaAspProLysAlaThr----- 940

DB 471 ACTACTACCACTACTACCAAGCTGACAGCAACAACTACTACTACCAAGCAGGA 412

QY 941 -----GluCysLysSerLeuGlnThrThrAspThrTyrValThr 953

DB 411 AGTGGAGAAAGATACAGAGAGATGATACATGATACAGAGAGATACATGCGTACG 352

QY 954 GlnThrSerThrHisThrSerThrSerThrIleThrSerThrIleThrSerLysIleThr 973

DB 351 AGCAGCTCATTCAGTACGAGTACGACAAAGTACGTCGACAGTACGTCGACAGTACG 292

QY 974 LeuThrSerThrArgArgCysLysProThrLysCysThrThr----- 987

DB 291 TTGACGTCGATGAGAAAGTGCAGCTTCCAAAGTATACACCGATTCCAAAGAGAGACA 232

QY 988 -----GlyGluGluAspAspAlaGlyAspValLysProSerGluGlyLeuArgMetSer 1005

DB 231 GATTAAGAGAGAGAGAGAGAGAGAGATGTTAAACCAATGATGGATGAATAAGA 172

QY 1006 GlyTyrAsnValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023

DB 171 GTTCTCATATGATTAATAATATGTTGCGAGAGATTGTTATGGCGATGATG 118

RESULT 13

AW332556/c

LOCUS AW332556 733 bp mRNA linear EST 31-JAN-2000

DEFINITION S9H7 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.

ACCESSION AW332556

VERSION AW332556.1 GI:6828913

KEYWORDS EST.

SOURCE Pneumocystis carinii

ORGANISM Pneumocystis carinii

Pneumocystis carinii

Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;

Pneumocystidaceae; Pneumocystis.

REFERENCE 1 (bases 1 to 733)

AUTHORS

Smolian, A.G., Arnold, J., Weise, M., Wunderlich, J., Staben, C.,
Edman, J.C., Kovacs, J. and Cushion, M.
Expressed sequence tags from *Pneumocystis carinii*
Unpublished (2000)

TITLE

Contact: Staben C

JOURNAL

School of Biological Sciences

COMMENT

University of Kentucky

CONTACT

101 Morgan Building, University of Kentucky, Lexington, KY

PUBLISHED

40506-0225, USA

UNPUBLISHED

Tel: 606 257 2161

FAX

Fax: 606 257 1717

E-MAIL

Email: Staben@pop.uky.edu.

FEATURES

Location/Qualifiers

ORIGIN

Alignment Scores:

Pred. No.:	4.68e-29	Length:	733
Score:	394.00	Matches:	84
Percent Similarity:	60.20%	Conservative:	37
Best Local Similarity:	41.79%	Mismatches:	46
Query Match:	7.28%	Indels:	34
DB:	2	Gaps:	6

US-10-654-416-14 (1-1023) x AW332556 (1-733)

QY 854 ValThrGluLeuGluAlaLysAlaPheAspLeuAlaGluValPheGlyArgTyrVal 873

DB 707 GTATCAAGACAGAGAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATG 648

QY 874 AspLeuLysGluArgCysAsnLysLeuGluSerAspCysArgIleLysGluAspCysLys 893

DB 647 GAATTTGAAGAGAAATCAATGCTTTGAACTAGATGGCTTTTAAAGAGATGTTGAG 588

QY 894 AspLeuGluGluValCysLysLysLeuAsnLysLeuGluSerAspCysArgIleLysGluAspCysLys 913

DB 587 GAATCTAAACAGCTTGTAAAGAAATGACACGTTATGCAAAAGAAATPAAACCATAGAA 528

QY 914 ValLysProHis-----GluThrValThrGluSerThrThrThrThrThrThrThr 928

DB 527 GTTACGCTTATCATACAGAGACATATATAAAGAAACCATGACATGATGATGATG 468

QY 929 -----ThrThrThrThrThrValAlaAsp 936

DB 467 ACAAAAGCATTTGACAAAGCTGATGGAAGACCGGTACGAAGACTATTTAGATGCCGAG 408

QY 937 Pro-----LysAlaThrGlu-----CysLysSerLeuGlnThrThrAspThrTyr 951

DB 407 TCACTAGGTGAGAGAAAGTAAACAGATGATCATTAAGTACGAGACAGATATATGG 348

QY 952 ValThrGlnThrSerThrHisThrSerThrSerThrIleThrSerThrIleThrSerLys 971

DB 347 GTACGAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 288

QY 972 IleThrLeuThrSerThrArgArgCysLysProThrLysCysThrThr----- 987

DB 287 GTACGTTGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 228

QY 988 -----GlyGluGluAspAspAlaGlyAspValLysProSerGluGlyLeu 1002

DB 227 GAGACAGATTAAGATG 177

Oy 1003 ArgMetSerGlyTTPaenValMetArgGlyValIleValAlaMetValIleSerPheMet 1022
 Db 176 AAAATTAAGATCTCGATATGATTAATAATGTTGGAGGATGATTTATGGGGATG 117
 Oy 1023 Ile 1023
 Db 116 ATG 114
 RESULT 14
 AM332531/c 710 bp mRNA linear EST 31-JAN-2000
 LOCUS Snp12 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
 ACCESSION AM332531
 VERSION AM332531.1 GI:6828888
 KEYWORDS EST.
 SOURCE Pneumocystis carinii
 ORGANISM Pneumocystis carinii
 Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
 Pneumocystidaceae; Pneumocystis.
 REFERENCE 1 (bases 1 to 710)
 AUTHORS Smilian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
 Edman,J.C., Kovacs,J. and Cushion,M.
 TITLE Expressed sequence tags from Pneumocystis carinii
 JOURNAL Unpublished (2000)
 COMMENT Contact: Staben C
 School of Biological Sciences
 University of Kentucky
 101 Morgan Building, University of Kentucky, Lexington, KY
 40506-0225, USA
 Tel: 606 257 2161
 Fax: 606 257 1717
 Email: staben@pop.uky.edu.
 FEATURES
 source
 1..710
 Location/Qualifiers
 /organism="Pneumocystis carinii"
 /mol_type="mRNA"
 /db_xref="taxon:4754"
 /lab_host="E. coli"
 /clone_lib="AGS-1"
 /note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
 P. carinii organisms (3x10e9) from a single rat (99-1-6,
 sacrificed on 3/17/99) at Cincinnati VA facilities.
 Trizol extracted RNA. Oligo dT priming, standard
 conditions described by vendor, Stratagene. Further
 details see www.uky.edu/Project/Pneumocystis/"
 ORIGIN
 Alignment Scores:
 Pred. No.: 5,62e-29 Length: 710
 Score: 393.00 Matches: 79
 Percent Similarity: 62.78% Conservative: 34
 Best Local Similarity: 43.89% Mismatches: 39
 Query Match: 7.26% Indels: 28
 DB: 2 Gaps: 5
 US-10-654-416-14 (1-1023) x AM332531 (1-710)
 Oy 872 TyrValAapLeuLysGluArgCysAsnLysLeuGluSerAapCysArgIleLysGluAap 891
 Db 691 TATTTGGAATTCGAAGAAGAGATGCAATGCTTAGAAGTGAATGGCGTTTAAAGAGAT 632
 Oy 892 CysLysAapLeuGluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysAap 911
 Db 631 TGTGAGGATCTAAACCAAGCTGTAAAGAAATAGAGAGATTGCAAGAAGATAGAAATCA 572
 Oy 912 LeuGluValLysProHisGluThr----- 919
 Db 571 TTTAAAGTTGGCGCCCTCATCTATACAGAGAAATATCAACCCCTACGGCGACC 512
 Oy 920 -----ValThrGluSerThrThrThrThrThrThrThrThrThrValAla 935
 Db 511 ACTTACGGCCACCTTTCAGACTACTACGAGCTACTACGACAACTACTACTACAAAC 452

Oy 936 AspPro-----LysAlaThr---GluCysLysSerLeuGluThrThrAspThr 950
 Db 451 AAGCGGGAAGTCGAGAGAAAAGTAAGAGAGTCAATGATAGACAGCCCAATACA 392
 Oy 951 TrpValThrGluThrSerThrHisThrSerThrThrIleThrSerThrIleThrSer 970
 Db 391 TGGGTGACGACGACGCTATTGCAATGCAAGTACCAACGACTACGCTACGCTACG 332
 Oy 971 LysIleThrLeuThrSerThrArgArgCysLysProThrLysCysThrThr-----Gly 988
 Db 331 ACACTGACGTTGACTTGATCGATCGCAGAGCTACCAAAATGATACTACTGATTAAC 272
 Oy 989 GluGluAapAapLysIle-----AspValLysProSerGluGlyLeuArg 1003
 Db 271 AGAAGACAGATTAAGAGAGAGAGAGAGAAAGATGTAAACCAATGATGAGGTGA 212
 Oy 1004 MetSerGlyTTPaenValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023
 Db 211 ATAAGAGTCTCGATATGATTAATAATGTTGGAGAGTATGTTATGGGATGATG 152
 RESULT 15
 AM334916/c 800 bp mRNA linear EST 31-JAN-2000
 LOCUS SAG03 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
 ACCESSION AM334916
 VERSION AM334916.1 GI:6831273
 KEYWORDS EST.
 SOURCE Pneumocystis carinii
 ORGANISM Pneumocystis carinii
 Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
 Pneumocystidaceae; Pneumocystis.
 REFERENCE 1 (bases 1 to 800)
 AUTHORS Smilian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
 Edman,J.C., Kovacs,J. and Cushion,M.
 TITLE Expressed sequence tags from Pneumocystis carinii
 JOURNAL Unpublished (2000)
 COMMENT Contact: Staben C
 School of Biological Sciences
 University of Kentucky
 101 Morgan Building, University of Kentucky, Lexington, KY
 40506-0225, USA
 Tel: 606 257 2161
 Fax: 606 257 1717
 Email: staben@pop.uky.edu.
 FEATURES
 source
 1..800
 Location/Qualifiers
 /organism="Pneumocystis carinii"
 /mol_type="mRNA"
 /db_xref="taxon:4754"
 /lab_host="E. coli"
 /clone_lib="AGS-1"
 /note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI;
 P. carinii organisms (3x10e9) from a single rat (99-1-6,
 sacrificed on 3/17/99) at Cincinnati VA facilities.
 Trizol extracted RNA. Oligo dT priming, standard
 conditions described by vendor, Stratagene. Further
 details see www.uky.edu/Project/Pneumocystis/"
 ORIGIN
 Alignment Scores:
 Pred. No.: 6.79e-29 Length: 800
 Score: 393.00 Matches: 84
 Percent Similarity: 47.14% Conservative: 48
 Best Local Similarity: 30.00% Mismatches: 122
 Query Match: 7.26% Indels: 26
 DB: 2 Gaps: 3
 US-10-654-416-14 (1-1023) x AM334916 (1-800)
 Oy 425 GluAsnAlaCysLysAapLysAlaArgAlaThrCysTyrLysArgGlyLeuAspAlaArgAla 444
 Db 784 GATCAAGGCTGTCAAAATGTACGACGCGCTATTAATAAGGACAAAGATAGAGATGTTG 725

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: October 29, 2005, 09:09:50 / Search time 396 Seconds
(without alignments)
4227.048 Million cell updates/sec

Title: US-10-654-416-14
Perfect score: 5411
Sequence: 1 MARAVRQAAAGTQNSIDEH.....MSGMWVGVVAVWVSPMI 1023

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_plus.p2n.model -DEV=xlh
-Q=/cg2_1/USFTO.epool/US1065416/runat_26102005_101024_21768/app_query.fasta_1.1155
-DB=Issued_Patents_NA -QPM=fastcap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=b1bs -START=1 -END=1 -MATRIX=b1bsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US1065416.qcch_1.1_237.qrunat_26102005_101024_21768 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NBS SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued_Patents_NA.*

- 1: /cg2_6/ptcdatata/1/ina/5A.COMB.seq.*
- 2: /cg2_6/ptcdatata/1/ina/5B.COMB.seq.*
- 3: /cg2_6/ptcdatata/1/ina/6A.COMB.seq.*
- 4: /cg2_6/ptcdatata/1/ina/6B.COMB.seq.*
- 5: /cg2_6/ptcdatata/1/ina/PCTUS.COMB.seq.*
- 6: /cg2_6/ptcdatata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5411	100.0	3072	4	US-09-762-724-13 Sequence 13, Appl
2	4398.5	81.3	3054	4	US-09-762-724-11 Sequence 11, Appl
3	4352	80.4	3084	4	US-09-762-724-7 Sequence 7, Appl
4	3827	70.7	3090	4	US-09-762-724-5 Sequence 5, Appl
5	3401.5	62.9	3081	4	US-09-762-724-9 Sequence 9, Appl
6	2762.5	51.1	3042	4	US-09-762-724-1 Sequence 1, Appl
7	2404.5	44.4	3006	4	US-09-762-724-3 Sequence 3, Appl
8	913	16.9	2814	1	US-07-781-034-1 Sequence 1, Appl
9	913	16.9	2814	5	PCT-US92-08328-1 Sequence 1, Appl
10	559	10.3	1448	1	US-07-781-034-3 Sequence 3, Appl
11	559	10.3	1448	5	PCT-US92-08328-3 Sequence 3, Appl
12	396	7.3	249	4	US-09-762-724-15 Sequence 15, Appl

13	274	5.1	6016	4	US-09-949-016-1054	Sequence 1054, Ap
14	274	5.1	6016	4	US-09-949-016-2264	Sequence 2264, Ap
15	262	4.8	4206	4	US-09-750-590A-3	Sequence 3, Appl
16	262	4.8	4730	4	US-09-750-590A-1	Sequence 1, Appl
17	261	4.8	5574	4	US-09-917-354-40	Sequence 40, Appl
18	261	4.8	6861	4	US-09-949-016-1240	Sequence 1240, Appl
19	261	4.8	6861	4	US-09-949-016-1241	Sequence 1241, Ap
20	261	4.8	6861	4	US-09-949-016-1242	Sequence 1242, Ap
21	260.5	4.8	6773	3	US-09-166-350-27	Sequence 27, Appl
22	260	4.8	5923	4	US-09-976-594-907	Sequence 907, Appl
23	253.5	4.7	6008	4	US-09-949-016-5058	Sequence 5058, Ap
24	253	4.7	5925	4	US-09-949-016-5233	Sequence 5233, Ap
25	252	4.7	6276	4	US-09-949-016-2009	Sequence 2009, Ap
26	252	4.7	6284	4	US-09-949-016-1028	Sequence 1028, Ap
27	252	4.7	10300	4	US-09-949-016-6-636	Sequence 636, App
28	251.5	4.6	5661	3	US-08-938-105-2	Sequence 2, Appl
29	251.5	4.6	7453	4	US-09-620-312D-248	Sequence 248, App
30	251.5	4.6	7501	4	US-09-620-312D-249	Sequence 249, App
31	251	4.6	8590	4	US-09-949-016-5562	Sequence 5562, Ap
32	251	4.6	38575	4	US-09-949-016-17304	Sequence 17304, A
33	251	4.6	119153	4	US-09-949-016-12378	Sequence 12378, A
34	247.5	4.6	6175	4	US-08-875-435B-1	Sequence 1, Appl
35	243.5	4.5	7596	4	US-09-023-655-1463	Sequence 1463, Ap
36	241	4.5	6644	4	US-08-875-435B-5	Sequence 5, Appl
37	240.5	4.4	6011	4	US-09-949-016-3017	Sequence 3017, Ap
38	237.5	4.4	10136	1	US-08-353-700-2	Sequence 2, Appl
39	237.5	4.4	10136	5	PCT-US95-16216-2	Sequence 2, Appl
40	237	4.4	9626	4	US-09-150-867-2	Sequence 2, Appl
41	232	4.3	8789	1	US-08-328-254-5	Sequence 5, Appl
42	231	4.3	4868	1	US-08-139-937-12	Sequence 12, Appl
43	231	4.3	4868	5	PCT-US93-11310-12	Sequence 12, Appl
44	231	4.3	5883	4	US-09-949-016-5001	Sequence 5001, Ap
45	228	4.2	4407	4	US-09-949-016-1690	Sequence 1690, Ap

ALIGNMENTS

RESULT 1
US-09-762-724-13
Sequence 13, Application US/09762724
Patent No. 6664053
GENERAL INFORMATION:
APPLICANT: Kovacs, et al.
TITLE OR INVENTION: Identification of a region of the major surface
FILE REFERENCE: 4239-58054
CURRENT APPLICATION NUMBER: US/09/762,724
CURRENT FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: PCT/US99/18750
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: US 60/096,805
PRIOR FILING DATE: 1998-08-17
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 3072
TYPE: DNA
ORGANISM: Pneumocystis carinii sp. f. hominis
NAME/KEY: CDS
LOCATION: (1) .. (3072)
US-09-762-724-13
Alignment Scores:
Pred. No.: 0
Score: 5411.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 4
Gaps: 0
Length: 3072
Matches: 1023
Conservative: 0
Mismatch: 0
Indels: 0

US-10-654-416-14 (1-1023) x US-09-762-724-13 (1-3072)

Qy 1 MetAlaArgAlaValLySAArgGlnAlaAlaGlyThrGlnAsnSerTLeaSPGluGluHis 20
Db 1 ATGGCGCGGGGCGTCAAGCGGACAGCAGGAGCACAGAAATGACATTGATGAGAGAACAT 60
Qy 21 ValLeuAlaLeuIleLeuLySGluAspGlyLeuSerGluGluGlnCysLySLyLeu 40
Db 61 GTTTTAGCTTTTAATTTCTAAAGGAGATGACTTAAGTGCAGGAATGCCAAAAAACTTA 120
Qy 41 LySLyTyrcysGlnGluLeuThrGlnAlaLySLeuaenileGluGlnValHisArgLyS 60
Db 121 AAAAAATTTGCCAAGAAATGACGTGAAGCAAAACCTAATATATAGAACAAAGTACACGAAAA 180
Qy 61 LeuLySGlyPheCysGluAspGlyLySLyAlaAspThrLySCysLySGluLeuLySLyAsn 80
Db 181 CTTAAAGGTTTTTGGCAAGATGAGAAAGCAGATACAAATATGCAAAAGAACTCAAAAGCCCAAT 240
Qy 81 IleGluLySLyCysThrThrIleLySGlySLySLyLeuLySGluAlaIleLySLySLyIle 100
Db 241 ATTGGAAAAAATGTACTACATCAAAAGAACTTAAGAAAGCAATTTAAAAAAAATTT 300
Qy 101 GlnIleIleThrAspLySAAspCysLySGluAsnGluGlnGlnCysLeuPheLeuGluGly 120
Db 301 CAGATTATTAAGGATTAAGATTGGCAAAAGAAATGAACAACATGCTATTTTGGAGGGA 360
Qy 121 ValCysSerLySGluLeuLySLyAspAspCysAsnThrLeuArgAsnLySCysTyrgLnuLyS 140
Db 361 GTATGTTCAAAAGAACTTAAAGATGATTCGAATCTTTGAGAAATTAAGTGCATCAAAAG 420
Qy 141 LySAArgAspLySLyValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGly 160
Db 421 AAACGTGATTAAGTTGCCGAAAGAGTTCTTTTAAGGACCTTCGTAGCGATCTTAATGGA 480
Qy 161 SerValIleCysGluLySLyLeuLySGluIleCysProValMetGlyArgGluSerAsp 180
Db 481 TCAGTCATATGTGAAAAAACTTAAAGAAATTTCCCTGTCATCGGAGGGAAGTAT 540
Qy 181 GluLeuThrAsnLeuCysLeuAsnGlnLySGluThrCysLySLyAsnIleLeuIleGluLyS 200
Db 541 GAGTTAAACAAACTGTGTCTGAACCAAGAAAGACATGTAAGAAATATTTTAAATGAAAAA 600
Qy 201 AspLySLyCysGlyThrLeuLySLyThrAspValSerAlaIleGluGlySerPheLySLyS 220
Db 601 GATTAAGAAAGTCGGTACTTTAAACAGATGTTTCAGACAGCTTAGAAAGTTTAAAAAA 660
Qy 221 GluThrCysLeuGluLeuLeuGluGlnCysTyrgPheTyrgIleLySLyAsnCysGlyAspAsp 240
Db 661 GAAACATGCTTGAATTAAGTCTGAACAAATGCTATTTTACATTGGAATATGCGAGAGACAC 720
Qy 241 AspIleIleIleLySCysIleGluLeuGlyGlyLySCysGlnGluGlnAsnIleAlaTyrmec 260
Db 721 GATTAATTAATTAATGTAATGAAATGGGAGAAATGCCCAAGAAACAAATATTTGCTTATATG 780
Qy 261 ProProGlyProAspPheAspProThrArgProGluAlaThrIleAlaGluAspIleGly 280
Db 781 CCACCAAGACCCGATTTTATTCACATGAGCCAGGCTTACAAATAGCAGAGATTAAGG 840
Qy 281 LeuGluGluPheTyrgLySLyValGluGluAspGlyValPheIleGlyLySLyAsnHisLeu 300
Db 841 CTGGAGAGCTTTTAAAGAGGTAGAGAGATGAGTTTATTTAGAAAGATCATCTA 900
Qy 301 ArgAspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLySLyLeuAsp 320
Db 901 AGAGATGCGACAGCTTTGTTGTCATTTGATCCAAAGATTCATAGCTTTAAAAAAAAC 960
Qy 321 AspLySGluLySCysGluGluAlaLeuGlnLySLySerCysLySLyAsnProHisGluHisGlu 340
Db 961 GACAAAGGAAATGCGAAGAGCCCTTCAAAAAGCTCAAAAATCTTCATGAACATAG 1020
Qy 341 AlaLeuGluSerLeuCysLySLySLyAsnGlyLeuSerAsnAspGlyThrLySLyCysGlu 360
Db 1021 GCTTTAGAAAGTTTAATGTAAGAAAAATGTGTTTAAGTATGATGGAACGAAAAAATGTCAA 1080
Qy 361 GluLeuGlnAsnAspIleAsnLySLyCysLySLyIlePheThrSerLySLyValThrAsnAsn 380

Db 1081 GAATTTGCAAAATGATTAATTAACAAACTTGCAAAATTTTCACTTCAAAAGTACATTAAT 1140
Qy 381 ArgLeuPheAspProThrLySLyAsnAsnGluIleValGlyTyrgGluGlyLeuProThr 400
Db 1141 CGTCTTTTGTATCCAAAGAAAGGAAATATATGAATTTGTTGATGCGAAAGGTTCCACAA 1200
Qy 401 PheLeuSerAsnGluAspCysAlaLySLyLeuGluSerTyrcysPheTyrgPheGlnLySLy 420
Db 1201 TTTCTTAGCAACGAAGATTGTGCCAAATTTGAGCTCATTTGTTTAAAAAAA 1260
Qy 421 CysProAspGlyGluAsnAlaCysLySLyAsnIleArgAlaThrCysTyrgLySLyLeu 440
Db 1261 TGTCACATGAGAAAAATGCAATGTAATAAGAACCAACATTTTCAAAAAGGAGCTT 1320
Qy 441 AspAlaArgAlaAsnLySLyValLeuGlnGluAsnMetArgGlyMetLeuHisGlySerAsn 460
Db 1321 GATTCACGGCGCAATTAAGTCTTCGCAAGAAATTTGCGAGGAATGTTACATGTTCAAC 1380
Qy 461 LySLySerTrpLeuGluLySLyPheGlnGluLeuValLySLyValCysGluLySLyLeuLySLy 480
Db 1381 AAAAGCTGGCTTGAAAGTTTCAACAAAGAAATTAAGTAAAGTATGATGCAAACTGAAAAA 1440
Qy 481 GluAsnLySGlySerPheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLyS 500
Db 1441 GAAAAACAAAGAGATTTCTCAAAAGATGAATTAATTTATGTTGTGTGACAGCCAGCAAA 1500
Qy 501 AlaAlaArgLeuLeuThrHisAspLeuArgMetLySLyThrIlePheLeuArgGlnIleu 520
Db 1501 GCAGCCCGGTTGCTTACACATGATCTTCGAATGAACCTATCTTTTACGACAAACACTG 1560
Qy 521 AspGluLySLyArgAspPheProThrAspLySLyAsnCysLySGluLeuGlyArgLySCysGln 540
Db 1561 GATCAAAAGCGAGATTTCCGACAGATTAATAAATTTGCAAGAAATGGGAGAAATGGCCAA 1620
Qy 541 AspLeuGlyGluAspSerLySLyIleThrTrpProCysHisThrLeuGluGlnGlyCys 560
Db 1621 GATTTAAGAGAGATTCAAAGAAATTAACATGATGATCTTACATGAGACACCAATGC 1680
Qy 561 AsnArgLeuGlyThrThrGlnIleLeuLySLyGlnValLeuLeuAspGluHisLySLyAspThr 580
Db 1681 AATGCTTTGGGACATCAGAAATTTTAAGCAGGTTTATTTGATGAACACAAAGATACT 1740
Qy 581 LeuLySLyAspGlnLySLyCysValLySLyTyrgLeuLySGluLySCysAsnLySLyTrpSerArg 600
Db 1741 TTGAAAGACCAAGAAATGTTGTGTAATACTTAATAAAGAAAGTGAATTAATGTTCTAGA 1800
Qy 601 ArgGlyAspAspArgPheSerPheValCysValPheGlnAsnAlaThrCysGluLeuMet 620
Db 1801 AGAGGAGATGACCGTTTCTTTTGTATGTCTTCCAAAACGCTACGTTGACCTGATG 1860
Qy 621 ValLySLyAspValLySLyAspArgCysGluValPheLySLyAsnIleLySLyAlaSerTyrgIle 640
Db 1861 GTAAAGACGGAAGACAGGTGTGAAGTATTCAAAAAATAATTAAGAGCTTCATATATT 1920
Qy 641 IleGluPheLeuGlnAsnAsnThrAsnLySLyIleThrThrLeuGluArgAsnCysProSer 660
Db 1921 ATTGAATTTCTTGAATAATTAATACAAATTAATAACAACTGGAAGAAATTTGCCCTCT 1980
Qy 661 TrpHisThrTyrcysAsnArgPheSerProAsnCysProGlyLeuThrLySLyGluAsnSer 680
Db 1981 TGGCATACGATTTGCAATAGATTTTTCATTAATTTGTCACAGGCTTTACGAAAGAAATAGT 2040
Qy 681 CysThrLySLyIleLySLyHisArgGluProPheTyrgLySLyAlaLeuGluAspAla 700
Db 2041 TGTAACAAATTAACAAAGACATGTAAGCCGTTCTATTAAGAAAGAGCCTTGAAGATGCT 2100
Qy 701 LeuLySLyValGluLeuGlnGlyLySLyLeuThrAspLySLySerLySCysGluProAlaLeuLyS 720
Db 2101 CTCAAGTAAGGCTTCAAGGAAATTTGATGATTAATCTAAATGTAACCTGCAATTTGAAA 2160
Qy 721 ArgTyrcysThrValAlaGlyAsnValAsnAsnAlaSerIleSerGlyLeuCysLySLyVal 740

```
Db 2161 AGATATTGTACAGTACGGGAAACGTAAATATGCTCAATCACTGCTTATGCAAACT 2220
Qy Aanthrilysaerpaenserglylyseraerpglubaepalaarglysergluleucyegluys 760
Db 741 AACCAGGATTAACCTCGGAAAGATGATAGATGCTTAAAGAAAGAACTTGTGAGAA 2280
Qy 2221 AACCCAGGATTAACCTCGGAAAGATGATAGATGCTTAAAGAAAGAACTTGTGAGAA 2280
Qy 761 LeuVallysergluValgluGluGlnCyelysaValaLeuprothrgluLeuGlyGlnProAla 780
Db 2281 TTACTGAAAGAGATGAGAAAGACATGCAACATTAACCAAGAAATTAAGCAACCCGCA 2340
Qy 781 AAlaerleuValylysaerpylylyethrlyrqluGluLeuValylysaValaGluGluAla 800
Db 2341 GCTGATCTTAAAAAAGATTAATAGCATATGAGAACTTAAAGAAACGGACGAGAGAAACA 2400
Qy 801 MetAenlyserSerleuValleuSerleuValleuValylysaerpaensergluys 820
Db 2401 ATGACAAAGTCAAGTCTTGTGCTCACTCACTTAAAGAAAGAAAGTAAATGATCAAAA 2460
Qy 821 SerAenSerlysaenlysaerpylysaValaValaSerAenglyleuGlnAerpyrThrlyls 840
Db 2461 AGTAATGCAAAAACAGGATTAAGATGCCGTTCAACGAGACCTTCAAGATACCAACAAA 2520
Qy 841 HlyVallylyleuValylysaerpylylyValylysaerpylylySerValThrGluLeuGluAla 860
Db 2521 CATGTGAAAAATCTACGAGGGGAGTTAAGGATGTATCCGTAAAGAAATTAAGAGCTAAA 2580
Qy 861 AAlaerpaerleuValaGluValaPheGlyArgrlyrValaLeuLeuValylysaerpylyls 880
Db 2581 GCATTGTATTTGGCAGCAGAAATTTGGAAAGATATGATGATTTGAAAGAAAGATGAT 2640
Qy 881 LylyleuGluSerAerpylylylyleuGluAerpylylysaerleuGluGluValaCyelyls 900
Db 2641 AAATTTGAGATCAGATTGAGAAATTAAGAGGATTTGCAAAAGACTTAAAGAAAGTATGCAA 2700
Qy 901 LylyleuValylysaerpylylysaerleuValylyleuGluValylyleuProthrgluThrVal 920
Db 2701 AAGATTAATAGGCTTGTGCAATCTGAAGCCTCTGAGGAGAAACCCGACGAAACAGTG 2760
Qy 921 ThrGlySerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 940
Db 2761 ACAGAAAGTACAGACACACACACACACACACACACACACACACACACACACACACACACAC 2820
Qy 941 GlnCyelylyserleuGlnThrThrAerpyrThrValThrGlnThrSerThrThrThrThrThr 960
Db 2821 GAATGCAAAATCTTACAGACACACACACACACACACACACACACACACACACACACACACAC 2880
Qy 961 ThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 980
Db 2881 ACCTCTACTATCATCATCATCATCATCAAAAATTAACATTAACATTAACATTAACATTAAC 2940
Qy 981 LylyleuProthrglylylyleuGluGluLeuAerpaerpylylyleuValylyleuProSerGlu 1000
Db 2941 AAACCAACCAAGTATACAGACAGAGAGAGATATGAGAGACGTAACCAAGTATGAG 3000
Qy 1001 GlylyleuAerpylylylyleuValylyleuValylyleuValylyleuValylyleuValyly 1020
Db 3001 GCGGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3060
Qy 1021 Pheneticle 1023
Db 3061 TTTCATGATT 3069
```

```
RESULT 2
US-09-762-724-11
; Sequence 11, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE REFERENCE: glycoprotein (MSG) gene of human Pneumocystis carinii
; CURRENT APPLICATION NUMBER: US/09/762.724
; CURRENT FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096, 805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 3054
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; NAME/KEY: CDS
; LOCATION: (1) .. (3054)
US-09-762-724-11
```

Alignment Scores:

Pred. No.:	0	Length:	3054
Score:	4398.50	Matches:	838
Percent Similarity:	88.81%	Conservative:	75
Best Local Similarity:	81.52%	Mismatches:	98
Query Match:	81.29%	Indels:	17
DB:	4	Gaps:	9

US-10-654-416-14 (1-1023) x US-09-762-724-11 (1-3054)

```
Qy 2 AAlaerpaerleuValylysaerpylylyleuGluValylyleuProthrgluThrVal 20
Db 1 GCGGGGCGGTCAACCGGACGTAACGACATCAACGCAATATGATGAAAGTGAAT 60
Qy 21 ValleuValleuValleuValleuValleuValleuValleuValleuValleuValleuValleu 40
Db 61 ATTTGGCGTTGATTTCTACAAAGAAATGCAATGGAATCAAAATGCAAAATGCAAAATGCAAA 120
Qy 41 LylylyerpyrlylyleuGluValleuValleuValleuValleuValleuValleuValleuValleu 60
Db 121 GAAATAATACGCGAAGGTTGAAAAAAGCACTTACATCAATGAAAAAGTACATTAATG 180
Qy 61 LeuValylylyleuGluValleuValleuValleuValleuValleuValleuValleuValleuValleu 77
Db 181 CTTAAAGATTTCTGTGAAATGGAATGGAAGCAAGTAACCAATTAACCAATTAACCAATTAACCA 240
Qy 78 LylyleuValleuValleuValleuValleuValleuValleuValleuValleuValleuValleuValleu 96
Db 241 CAAGCCAAAGTTACGGGGAATGTACAAATTTTAAACAAAGATTAAGACCAAGCTTA 300
Qy 97 LylylyerpylylyleuValleuValleuValleuValleuValleuValleuValleuValleuValleu 116
Db 301 ACAAAATCA-----TCAGATGATTAATGCAAAAGAGTGAACGACCAATGCTTA 348
Qy 117 PheleuGluGlyValaCyserlylyleuValylyleuValylyleuValylyleuValylyleuValyly 136
Db 349 TTTTGGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 405
Qy 137 CyelylyleuValylylyleuValylylyleuValylylyleuValylylyleuValylylyleuValyly 156
Db 406 TGTTACCAAGAAAAAGCTGACGAGTAGCAAGAAAGTCTTTGAGGCGCCTTGTAGT 465
Qy 157 AspleuValleuValleuValleuValleuValleuValleuValleuValleuValleuValleuValleu 176
Db 466 GATCTCAATTAACCAAGAAACATGAAAAAACTGAAAAAGATTTCCCAAGCTTTCAG 525
Qy 177 ArgGlySerAerpylylyleuValleuValleuValleuValleuValleuValleuValleuValleuValleu 196
Db 526 AGGAAAGTAAATTAACGACGATTTGTTGAAACCAAGAAAAAGATGCGAGAAATTT 585
Qy 197 LeuValleuValleuValleuValleuValleuValleuValleuValleuValleuValleuValleuValleu 216
Db 586 ATAAAGAAAAAGTAAATTAATGCACTTAAACCAATGTTGCAACAGCACTTGA 645
Qy 217 SerPheleuValylylyleuValleuValleuValleuValleuValleuValleuValleuValleuValleu 236
Db 646 AGTTTAAAAAAGAAATATGCTTGAATTACTTGAAACAAATGCTAATTTTACATTTGAAAT 705
```

[illegible]

Db	1774	AAATAATGGCTCTGAAGAGAGATGACCGCTTCTCTTTGTATGCTCTTCCAAACGCT	1833
Oy	616	ThrCysGluUueCValLysAspValLysAspArgCysGluValPheLysLysAsnIle	635
Db	1834	ACGTGTAGCTGATGCTAAAGACGTGAAGACAGCGGTGMAATTATTCAAAAAAATATA	1893
Oy	636	LysAlaSerYrIleIleGluPheLeuGluAsnAsnThrAsnLysIleThrThrLeuGlu	655
Db	1894	AAACCTTCATATATATTAATTAATTTCTTGAATAATATACAAATATAAACACACCTGGAA	1953
Oy	656	ArgAsnCysProSerTPH1sThrYrCysAsnArgPheSerProAsnCysProGluLeu	675
Db	1954	AGAAATGTCCTCTTGGCATAGCTATTCGAAATAGATTTTCACTTAATTTGTCAGGCTT	2013
Oy	676	ThrLysGluAsnSerCysThrLysIleLysLysHisArgGluProPheYrLysArgLys	695
Db	2014	ACGAAAGAAGATAGTTGTACAAAATACAGACATTTGTAGCCGCTTCTATAAAGAAAG	2073
Oy	696	AlaLeuGluAspAlaLeuLysValGluLeuGlnGlyLysLeuThrAspLysSerLysCys	715
Db	2074	GCCTTGGAAGATGCTCTCAAAAGTAGACCTTCAAGGAAAAATGACTGATTAATCTAAATGT	2133
Oy	716	GluProAlaLeuLysAspYrCysThrValAlaGlyAsnValAsnAsnAlaSerIleSer	735
Db	2134	GAACCTTCATTTGAATAATATTTGTACATGACGGGAACCTTAATATATGCGTCAATAGT	2193
Oy	736	GlyLeuCysLysAlaAsnThrLysAspAsnSerGlyLysSerAspGluAspAlaArgLys	755
Db	2194	GGCTTATGCAAAAGTTACACCAAGATATCTTGAAAGAGGTGATGAGATGCTAGAAAG	2253
Oy	756	GluLeuCysGluLysLeuValLysGluValGluGluGlnCysLysAlaLeuProThrGlu	775
Db	2254	GAACCTCTGTAGAAATTAAGTGAAGAAAGTGAAGACAGTGCACAAAGCATTAACCAACGAA	2313
Oy	776	LeuGlyGlnProAlaAlaAspLeuLysAspDyrLysThrYrGluGluLysLys	795
Db	2314	TTAGAGACACCGGAGCTGATTTTAAAAAAGATTATTAAGCATATGAGAACTTAACAAA	2373
Oy	796	ArgAlaGluGluAlaMetAsnLysSerSerLeuValLeuSerLeuIleLysLysAsnGlu	815
Db	2374	CGTCCAGAGAGACAAAGAACAGACTCCAGCTCTGTTGTTCACATCATTAAAGAAAAACGAA	2433
Oy	816	SerAsnValSerLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGlyLeu	835
Db	2434	AGTATGTATCAAAAAAGTAAATAGCAAAAAACAGGATTAAGATGCCGTTTCAAAACGGACTT	2493
Oy	836	GlnAspThrThrLysIleLysLysIleLeuAspArgGlyValLysAspValSerValThr	855
Db	2494	CAAAATATCCACAAACATGTGAAAAATCTACGGAGAGAGATTAAGATGATTCCTGTACA	2553
Oy	856	GluLeuGluAlaLysAlaPheAspLeuAlaAlaGluValPheGlyArgYrValAspLeu	875
Db	2554	GAATTAAGACTTAAGACATTTGATTTGGACGACGAAATATTGGAGATATGTRGATTTG	2613
Oy	876	LysGluArgCysAsnLysLeuGluSerAspCysArgIleLysGluAspCysLysAspLeu	895
Db	2614	AAAGAAAGATGATAATTAATTCGAAATCGAATTCGCAATTTAAGAGGAGATTCGCAAAAGCTTA	2673
Oy	896	GluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluValLys	915
Db	2674	GAAGAAATATGCAAAAAAGATTAATTAAGCTGTGTCGCAATTCGAAGCCTTCGGAGGTGAAG	2733
Oy	916	ProHisGluThrValThrGluSerThrThrThrThrThrThrThrThrThrValAla	935
Db	2734	CCGACAGAAACGTGACAGAAAGTACACAGCAATACACACACACACACACACCCGTTGCC	2793
Oy	936	AspProLysAlaThrGluCysLysSerLeuGlnThrThrAspThrTrpValThrGlnThr	955
Db	2794	GATCCGAAGGCAACGGAATGCAATCCTTACAGACAAACAGACATCGGTTTACACAGACA	2853
Oy	956	SerThrHisThrSerThrSerThrIleThrSerThrIleThrSerLysIleThrLeuThr	975

QY 466 LysPheGInGInGluLeuValLysValCySGluLysLeuLysLysGluLysSer 485
 Db 1417 GAGTTTCAACAAAGATTGTAAGTAATGTAAGGACTA---AAAGAAATAAAGAAAT 1473
 QY 486 PheSerAAspGluLeuPheIleLeuCyValGlnProAlaLysAlaIAspLeu 505
 Db 1474 TTCCAAACGATGAATATTTCTCTGTGTACGACGCAAAAGCTGCACGATTACTT 1533
 QY 506 ThrHisAAspLeuArgMetLysThrIlePheLeuArgGInGInLeuAAspGlnLysArgAsp 525
 Db 1534 ACACACGATCATCAATAGAGGATTACCTTTTACGACAAACAATTGATCAAAAGAGAGAT 1593
 QY 526 PheProThrAAspLysAAspCyLysGluLeuGlyAlaGlyLysCySGlnAAspLeuGlyLysP 545
 Db 1594 TTTCCGACAGATTAAGGACTGCAAGGAACTAGGGAATAATGCAAGATTGGAAGAAAGAT 1653
 QY 546 SerLysGluIleThrTriProCyHisThrLeuGInGInGlnCyAAspArgLeuGlyThr 565
 Db 1654 TCAGAAAGAAATTAACATGCGCATGTCTACATCGAGCGCAATGCAATCGCTTGGGACT 1713
 QY 566 ThrGluIleLeuLysGlnValLeuLeuAAspGluHisLysAAspThrLeuLysAAspGlnL 585
 Db 1714 ACAGAAATTTTAAACAGATTATTGATGTAACACAAGATACCTTGAAAGACCAAGAA 1773
 QY 586 SerCySValLysTyrLeuLysGluLysCySAsnLysTyrSerArgArgGlyLysAAspArg 605
 Db 1774 AGTTGTGTAATAATACCTTAAGAAAGAAAGTAAATAATGCTTAAGAAAGAGATACCGT 1833
 QY 606 PheSerPheValCyValPheGlnAAsnAlaThrCySGluLeuMetValLysAAspValLys 625
 Db 1834 TTCTCTTTTGTATGTCTTTTCCAAACCGTACGTGTGACTGATGTAAGACCTGAAA 1893
 QY 626 AAspArgCySGluValPheLysLysAAsnLysLysAlaSerTyrIleIleGluPheLeuGlu 645
 Db 1894 GACAGCTGTGAAGTATTCAAAAAAATATTAAGCTTCATATATTGAAATTTCTTGA 1953
 QY 646 AAsnAntThrAAsnLysIleThrThrLeuGluArgAAspProSerThrHisThrTyrCyS 665
 Db 1954 AATAATACAAATAAATAAACAACACTGAAAGAAATTTGCTTGGCATACGTAATGCTC 2013
 QY 666 AAsnArgPheSerProAAspCySProGlyLeuThrLysGluAAspSerCySThrLysIleLys 685
 Db 2014 AATATATTTTCCCTTAATTTGTCACAGCTTACGAAAGAAATGTTGTACAAATTCAG 2073
 QY 686 LysHisArgGluProPheTyrLysArgLysAlaLeuGluAAspAlaLeuLysValGluLeu 705
 Db 2074 AAGCATTTGTGAGCCGCTTCTATTAAGAAAGCCCTTGAAAGATGCTCTCAAGTAGAGCTT 2133
 QY 706 GInGlyLysLeuThrAAspLysSerLysCySGluProAlaLeuLysArgTyrCySThrVal 725
 Db 2134 CAAGGAAATTAATGATTAATCTAAATGTGAACCTGCAATTGAAACAGATATTGTACAGTA 2193
 QY 726 AlaGlyAAsnValAAsnAAsnAlaSerIleSerGlyLeuCySValAAsnThrLysAAspAsn 745
 Db 2194 GCGGGAACGTTAAATATATCCGCAATCGACTGCTTATGCAAAAGCTTAACCAAGATTAAC 2253
 QY 746 SerGlyLysSerAAspGluAAspAlaArgLysGluLeuCySGluLysLeuValLysGluVal 765
 Db 2254 TCTGGAAGAGATGAGAGATGCTAGAAAGAACTCTGTGAGAAATTCAGTGAAGAAAGTG 2313
 QY 766 GInGluGInCySValAlaLeuProThrGluLeuGInGInProAlaAlaAAspLeuLysLys 785
 Db 2314 GAAGGAACGATGCAAGATTAACCAAGAAATTAAGCAACCGGCGAGCTGATCTAAAAAAA 2373
 QY 786 AAspTyrLysThrTyrGluGluLeuLysLysArgAlaGluGluAlaMetAAsnLysSerSer 805
 Db 2374 GATTATTAAGACATTAAGAGAACTTAAGAAACCTGCAGAGAAACAAATGAACAGTCCAGT 2433
 QY 806 LeuValLysSerLeuIleLysLysArgGlnSerAAsnValSerLysSerAAspSerLysAsn 825
 Db 2434 CTGTGTTTGTCACTCATTAAGAAAAACGAAGTAATGTATCAAAAAGTAATGCAAAAAC 2493
 QY 826 LysAAspLysAAsnAlaValSerAAsnGlyLeuGlnAAspThrThrLysHisValLysIleLeu 845

Db 2494 AAGATTAAGAAATGCGTTTCAACGAGCTTCAGATACCAACAAACATGTGAAATTACTA 2553
 QY 846 ArgArgGlyValLysAAspValSerValThrGluLeuGluAlaLysAlaAAspLeuAla 865
 Db 2554 CGGAGAGAGATTAAGATGTATCCGTAAACAGAAATTAGAGCTTAAGATTTGATTGGCA 2613
 QY 866 AlaGluValPheGlyArgTyrValAAspLeuLysGluArgCySAsnLysLysGluSerAsp 885
 Db 2614 GCAGAGATATTGGAGAAATGTAGATTGGAAGAAAGATGTAAATTAATGGAAATCGAGAT 2673
 QY 886 CySArgLysLeuLysAAspCySValAAspLeuGluGluValCySValLysIleAAsnLysAla 905
 Db 2674 TCAGAAATTAAGAGGATTTGCAAGACTTACAGAAAGTATGCAAAAGATTAAATTAAGCTT 2733
 QY 906 CySArgAAsnLysProLeuGluValLysProHisGluThrValThrGluSerThrThr 925
 Db 2734 TGTGCAATCTGAAGCCCTCGAGAGTGAAAGCCGACGAACAGTGAACGAAGTACAAAG 2793
 QY 926 ThrThrThrThrThrThrThrThrValAlaAAspProLysAlaThrGluCySValSerLeu 945
 Db 2794 ACAACTACAAACAAACAAACAAACCGTTGCCGATCCGAAGGCAACGGAATGCAAAATCCCTTA 2853
 QY 946 GlnThrThrAAspThrThrProValThrGlnThrSerThrHisThrSerThrSerThrIleThr 965
 Db 2854 CAGACAAACAGACATGAGGTACACAGACATGCACACACACACACAGCTTACTATACCA 2913
 QY 966 SerThrIleThrSerLysIleThrLeuThrSerThrArgArgCySValProThrLysCyS 985
 Db 2914 TCTACATCATCATCAAAATAATTAACATTGACATCAACGAGGCGATCAAAACCAACAGTGT 2973
 QY 986 ThrThrGlyGluGluAAspAAspAlaGlyAAspValLysProSerGluGluLysArgMetSer 1005
 Db 2974 ACAGACAGG-----GATGATGCAGAAAGACGTGAAGCCAACTGAAGCTTGAGGGGTGAGC 3027
 QY 1006 GlyTrpAAsnValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023
 Db 3028 GGGTGAATGTGATGAGGGGGGTGATAGTACCAATGTTATTTCCTTCATGATTT 3081

RESULT 4
 US-09-762-724-5
 ; Sequence 5, Application US/09762724
 ; Patent No. 6664053
 ; GENERAL INFORMATION:
 ; APPLICANT: Kovacs, et al.
 ; TITLE OF INVENTION: Identification of a region of the major surface
 ; FILE REFERENCE: 4239-58054
 ; CURRENT APPLICATION NUMBER: US/09/762,724
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: PCT/US99/18750
 ; PRIOR FILING DATE: 1999-08-17
 ; PRIOR FILING DATE: 1998-08-17
 ; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5
 ; LENGTH: 3090
 ; TYPE: DNA
 ; ORGANISM: Pneumocystis carinii sp. f. hominis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(3090)
 US-09-762-724-5

Alignment Scores:
 Pred. No.: 0 Length: 3090
 Score: 3827.00 Matches: 728
 Percent Similarity: 81.54% Conservative: 120
 Best Local Similarity: 70.00% Mismatches: 164
 Query Match: 70.73% Indels: 28
 DB: 4 Gaps: 8

QY	705	LeuGInGlyLysLeuThrAspLysSerIleCysGluProAlaLeuLysArgTYrCysThr	724
Db	2134	CTTCGAGGAAATTTAAGTAATTAATCTAAAGTGAACCTGCATTGAAAGATATTGTACA	21933
QY	725	ValAlaGlyAenValAsnAsnAlaSerIleSerGlyLeuCysLysAlaAsnThrLysAsp	744
Db	2194	GTAATTGAAGAAGCTTAATTAATGCGTCATCGACAGCTTATTATGTAAGAATATACGAAAGT	22533
QY	745	AsnSerGlyLysSerAsp--GluAspAlaArgLysGluLeuCysGluLysLeuValLys	763
Db	2254	AAAACTAAAAAGCCGATATATAAAATGTTAGAAAGAAAGCTTGTCTAAATTAGTGGAA	23133
QY	764	GluValGluGluGlnCysLeuAlaLeuProThrGlyLeuGlnProAlaAlaAspLeu	783
Db	2314	GAGGTGAACACGACATGCAAGATATTACCAACAGATTTACACAGCTCGAAAAAGTCTA	23733
QY	784	LysLysAspTYrLysThrTYrGlyGluLeuLysLysArgAlaGluGluAlaMetAsnLys	803
Db	2374	AAAAAAGATGTTAAGACATATAGAGAACTTTAAGAAAGGCAAAAAAGCAATGAACAG	24333
QY	804	SerSerLeuValLeuSerLeuIleLysLysAsnGluSerAenValSerLysSerAsnSer	823
Db	2434	TCCAGCCTTGTTTATCACTGTTTAAAGAAAAAGAAAGTAATATACATCGAAAAATATATAGC	24933
QY	824	LysAsnLysAspLysAsnAlaValSerAsnGlyLeuGlnAspThrThrLysHisValLys	843
Db	2494	AAAAACAAAGATTAAGAAATGTCGTTCAACGACCTTCAAGATATCCACAAAATATGTGTAA	25533
QY	844	IleLeuArgArgGlyValLysAspValSerValThrGluLeuGluValLysAlaPheAsp	863
Db	2554	ATACTTCAGAAAGAGTTAAGAGGACCTTTACACAAATCTGAAGCCAAAGCATTTGAT	26133
QY	864	LeuAlaAlaGluValPheGlyArgTYrValAspLeuLysGluArgCysAsnLysLeuGlu	883
Db	2614	TTGGCAGCAGAAAGTGTGGAAGATATGACCTTGAAAGAAAAATGTGAGAAATTGACT	26733
QY	884	SerAspCysArgIleLysGluAspCysLysAspLeuGluGluValCysLysLysIleAsn	903
Db	2674	TCCGATTCGGGATTTAAAGACATTCGATGGTTTAAAAAGAGTGTGGAAGAGTTGAG	27333
QY	904	LysAlaCysArgAsnLeuLysProLeuGluValLysProHisGluThrValThrGluSer	923
Db	2734	AAGACATGTCACGATCTTAAGCCTTGAAGGTGAAGTGCATGAATATGTACACAGAAC	27933
QY	924	ThrThrThrThrThrThrThrThrThrValAlaAspProLysAlaThrGluCysLys	943
Db	2794	ACAACGACGACCAACAGACAAACGACCGTTACCGATCCGAAGGCAACAGAAATGCAA	28533
QY	944	SerLeuGlnThrThrAspThrTYrValThrGlnThrSerThrHisThrSerThrSerThr	963
Db	2854	TCCATTACGAAACAGATACATGGGTTTACACAGACATGACACACACAGACCGCTTACC	29133
QY	964	IleThrSerThrIleThrSerLysIleThrLeuThrSerThrArgArgCysLysProThr	983
Db	2914	ATCAATATTACATCAATCAAAAAATTAATTGATCAACAGGCGATGCAAAACCAACC	29733
QY	984	LysCysThrThrGlyGluGluAspAspAlaGlyAspValLysProSerGluGlyLeuArg	10033
Db	2974	AAAGTTACGACACGGG-----GATGAACAGGAGACGTGAACCAAGTGAAGGATTGAAG	30273
QY	1004	MetSerGlyTYrPheAenValMetArgGlyValIleValAlaMetValIleSerPheMetIle	10233
Db	3028	ATGAATGGGTGAGCGATGATGAGGGGGGTGATAGTACAAATGGTTATTTCCGTCATGATT	30873

[illegible]

[illegible]

Db 1300 AATATCAGACGACATGCTTACAGAGGACCTTACACATTAGCAAAATGATATTCACA 1359
 Qy GluabmefaraglymetleuHIGlySeranlyserTrpleuGluLyPheGln 469
 Db 1360 AAGAAATGCGAGAAATGCTCAATGTTCAAAATACATGCTTACGTTCCAAAA 1419
 Qy 470 GluLeuValyValCyegLulyleuLyblysgLubnlysgLyserPheSerAsnAsp 489
 Db 1420 AAATCTATAGAAAGTGTGCAAAAAGTGAAGAAATGAAGATTAAGAGTTTCCAGTAAT 1479
 Qy 490 GluLeuPheHileuCyevaGlnProAlalyalaaIaaleuLeuThiAspLeu 509
 Db 1480 GAATTAATTTGCTATGTGTCAACCAATCAAAAGACGCTCATGCTTCCATGATCTT 1539
 Qy 510 ArgmetLeuThriIepheleuHxGlnGlnleuAspGlnLybAspAspPheProThAsp 529
 Db 1540 CCGATGAAATCTATCTTTTTCAGAAAGCATTTGAAACGAAAGCCAGATTTTCCAGTAAA 1599
 Qy 530 LybAsnCyblysgLuleuGlyArgLybCyegLlnAspLeuGlyAspSerLybGluLe 549
 Db 1600 GAAGATGCGAGAAATTAATTAAGAAATGTGAGGCTCAAGAAAGATTTCAAAAAATT 1659
 Qy 550 ThrTrProCybHsthIreuleuGlnGlnCybAsnArgleuGlyThrGluIleu 569
 Db 1660 GAATGGCCATGTCAATCAATTAAGCCAAATTTGTGATCAATTAAGAAACGCTAAAGAAATTG 1719
 Qy 570 LybGlnValleuLeuAspGlnHileuAspThriIeuleuAspGlnIuserCyevaLyb 589
 Db 1720 AAAAACTTTTAACTTAATGAACATTAAGATTAATTTGAAATTAAGAAATTTGGAATG 1779
 Qy 590 TytleuLybGluLyCybAsnLybTrpSerArgLybGlyAspAspAspPheSerPheVal 609
 Db 1780 TATTTGAAGAGAAATGAATGAATGCTTGAAGAGAAATGAACGTTTCTCTTTTA 1839
 Qy 610 CyevaIlePheGlnAsnAlaThrCyegLuleuMetValyAspValLybAspArgCyegLu 629
 Db 1840 TGTCCTTTGCAAAATAGACCTTGCAAGATTAATGTGAGAAAGATGTGAAAAATCAATGCAAA 1899
 Qy 630 ValIleuLybAsnIleuLybAlaSerTrileuGluPheleuGluAsnAsnThAsn 649
 Db 1900 ATATTTGAAAAAACATTAATAAAATACCAAGCATATGAT-----AGTAAACT 1947
 Qy 650 LybIleThrThreuleuGluArgAsnCybProSerTrpHsthThrLyCybAsnArgPheSer 669
 Db 1948 ABAATTAAGAAATTAAGGACATATTTGCTTATTTGGCAACCACTGSCATAGATTGGA 2007
 Qy 670 ProAsnCybProGlyLeuThrLybGluAsnSerCybThrLybIleuLybHileuGlu 689
 Db 2008 CCCAATTTGCCCGGATCTTGAAAAA--AAATAATGTGAAGACTTTGAAAAAATATTGCAAA 2064
 Qy 690 ProPheTrlybAsglybAlaIleuGluAspAlaIleuLybValGluIleuGlnLybIleu 709
 Db 2065 CCTTAATTAAGAAAGACCTTGAAATGACCTTAATTAATTTGAGTTTGAAGGACATCTT 2124
 Qy 710 ThrAspLybSerLyCyegLulProAlaIleuLybAspArgLybCybTrpValAlaGlyAsnVal 729
 Db 2125 GATTAAGAAAAAACTGCAAAACAAATCTTGATAAGTACTGTAACATTAAGGATCAAAACA 2184
 Qy 730 AsnAsnAlaSerIleSerGlyLeuCybLybAlaAsnThrLybAspAsnSerGlyLybSer 749
 Db 2185 GGAATTAATAAACCTTAAGGTTTGTGTAACAGTTCTACTGATTAACAAT----- 2232
 Qy 750 AspGluAspAlaArgLybGluLeuCyegLulLybLeuValLybGluValGluGlnCyb 769
 Db 2233 ---GAAACATTTAAGATTAACCTTTGCAAAAACATGTTCAAGCGGTGAAAGAAAAAGC 2289
 Qy 770 LybAlaIleuProThGlnLeuGlyGlnProAlaAlaIleuPheLybLybAspTrpLybThr 789
 Db 2290 CAAGATTAATCAAAAGAACTTGAAGAAACAAATGATTTTGAAGAAACAAATTAAGAT 2349
 Qy 790 TyrgLulLeuLybLybArgAlaGluIuAlaMetAsnLybSerSerLeuValIleuSer 809
 Db 2350 TATGAAAAAGTAAAAAGATTAACAAAAATGCAATGAAGAAACAAATCTCGTTTTTTCA 2409

Qy 810 LeuIleLybAsnGluSerAsnValSerLybSerAsnSerLybAsnLybAspLybAsn 829
 Db 2410 ACACTTAATCAACAGATTAATAAAGAAAAAGG-----GTCAGACCT 2454
 Qy 830 AlaValSerAsnGlyLeuGlnAspThrThrLybHileuValyIleuAlaGlyGlyVal 849
 Db 2455 AGTAAGCCTGTGTAGTCAAGATATTGTA---CATTTTAACCTGTGTAAGAAAGATGA 2511
 Qy 850 LybAspValSerValThrGluLeuGluAlaIleuAlaPheAspLeuAlaIleuValPhe 869
 Db 2512 AAA---GTTCAAGTGAAGAAAAAGGCAAAAGCGTTTATTTGTAAGCATTAAGCATTC 2568
 Qy 870 GlyArgTrValAspLeuLybGluArgCybAsnLybIleuSerAspCybArgIleLyb 889
 Db 2569 AGCTTTATGTAGAGTTAAAGAAAGCGTGCACATCTAAAGATGATTCGAAATTTAGA 2628
 Qy 890 GluAspCybLybAspLeuGluGluValCybLybLybIleAsnLybAlaCybAspAsnLeu 909
 Db 2629 AAAGAAATGTAAA---TGTAAGACCAAGTGCAGAAAGATAGAAAAAATATGTTAAAAATA 2685
 Qy 910 LybProleuGluValyAspProHileuThrValIleuThrGluSerThrThrThrThr 929
 Db 2686 GAACCACTGAAGTAAAGCCACATGAATTAATAACGTAACGGAACCAACATTAACACA 2745
 Qy 930 ThrThrThrValAlaAspProLybAlaThr-----GluCybLybSer 944
 Db 2746 GTCAAGAAACAGTCMAAGAAAGCAAGAAAAACAGTAGAGACGAGAAATGCAATCT 2805
 Qy 945 LeuGlnThrThrAspThrTrpValThrGlnThrSerThriHsthSerThrThrIle 964
 Db 2806 CTCAGCAACAACAGACAGTGGTCACAAAAGCCTCAACCCATACAGACACCTCCACACT 2865
 Qy 965 ThreSerThriIethSerLybIleThreuleuThreSerThrArgArgCybLybProThLyb 984
 Db 2866 ACGTCCACAGTTCAGTCMAAGATTAACCTGACGAGGCGGTGTAAAGCTTAAGCGAAG 2925
 Qy 985 CyethThrGlyGluGluAspAspAlaGlyAlaAspValyAspProSerGluGlyLeuArgMet 1004
 Db 2926 TGTACGACAGAGAGAGAAATGAAGACGAGAGAGTGAAGCGAGTGAAGGCGGTGAGATG 2985
 Qy 1005 SerGlyTrpAsnValMetArgGlyValIleValAlaMetValIleSerPheMetIle 1023
 Db 2986 AGTGGGTGAGTGTGATGAAGAGGGGTGTATTAGCAATGATGATTCATTCATGATT 3042
 RESULT 7
 US-09-762-724-3
 ; Sequence 3, Application US/09762724
 ; Patent No. 6664053
 ; GENERAL INFORMATION:
 ; APPLICANT: Kovacs, et al.
 ; TITLE OF INVENTION: Identification of a region of the major surface
 ; FILE REFERENCE: 4239-58054
 ; CURRENT APPLICATION NUMBER: US/09/762, 724
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: PCT/US99/18750
 ; PRIOR FILING DATE: 1999-08-17
 ; PRIOR APPLICATION NUMBER: US 60/096, 805
 ; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 3006
 ; TYPE: DNA
 ; ORGANISM: Pneumocystis carinii sp. f. hominis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(3006)
 US-09-762-724-3
 Alignment Scores: 5.37e-251 Length: 3006
 Pred. No.:


```

Db      1981 GGAGGAAAAGATGGCAATGCAAAAACTCAATATAAAGTCAAATCATCTTCTGAAAAA 2040
Qy      695 LysAlaLeuGluAaPAlaLeuValGluLeuGlnGlyLysLeuThrAaPLeuSerLys 714
Db      2041 GAACTTTAGAAAATTAAGATTGTGAAGAAATTTGAAAGGTAAGTATTAACAACGTAGAGAA 2100
Qy      715 CysGluProAlaLeuLysArgTyrCysThrValAlaGlyAaPValAaPAlaSerLys 734
Db      2101 TGTAAACAATACCTTAATATATATACGTACACAATGAAAAAGCAGAAAT--GGGTG 2157
Qy      735 SerGlyLeuCybLysValAaPThrLysAaPAsnSerGlyLysSerArgGluAaPAlaArg 754
Db      2158 GAAACTTTGGTCCAAA-----AGCAAAAGAAAACACC--AAGAGTGCATTAAAGTTAGA 2208
Qy      755 LysGluLeuCybGlyLysLeuValGluValGluGlnGlnGlyLysValAaPLeuProThr 774
Db      2209 GAAAGACTCTGTGAAAAGCTTAATTAACGTATTAAGAAAATGCTCAAAATTAAGAGAC 2268
Qy      775 GluLeuGlyGlnProAlaAaPLeuLysLysAaPValAaPValAaPValAaPValAaPVal 794
Db      2269 GAGCTTGAAGAAATTAAGAGGCTTTGAAAAGAAAAGAAAGAAAGTAAATAAATAAATAA 2328
Qy      795 LysArgAlaGluGluAlaMetAaPLeuSerSerLeuValLeuSerLeuLysLysAaPVal 814
Db      2329 GAAAGAACAGAAAAGCCATGGAAGATGCAAAACCTTATTTATCGAAGCGAAAGACCT 2388
Qy      815 GluSerAaPValSerLysSerAaPValSerLysAaPValAaPValAaPValAaPValAaPVal 834
Db      2389 GATTAATTAATTAATTAAG-----TCAGTAATTAAGCTCATCTGATACCTTAAGAAA 2442
Qy      835 LeuGlnAaPThrThrLysValAaPValLysValAaPValAaPValAaPValAaPValAaPVal 854
Db      2443 GGAAGAAAGCAACAAC-----GATTTAACTTGTAAAGAAATGCAAAA--GTCCAGCTA 2496
Qy      855 ThrGluLeuGluAlaLysAaPLeuAaPLeuAaPLeuAaPLeuAaPLeuAaPValAaPVal 874
Db      2497 ACAGAAAAGAAATTAAGACGATTTGATTTGATGCAAGACGATTTGATTTGATTTGATTTG 2556
Qy      875 LeuLysGluAaPValAaPValAaPValAaPValAaPValAaPValAaPValAaPValAaPVal 894
Db      2557 TTGAAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2613
Qy      895 LeuGluGluAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 914
Db      2614 TGTGAGAGATCCATTAATAAAGATACAGGAAATATGTTCAACATTAAGAGCCACTAAAGTG 2673
Qy      915 LysProHISGluThrValThrGlu-----SerThrThrThrThrThrThrThrThrThr 932
Db      2674 AGACCAACAGAAATTAAGTAAATAAACAATTAACAATTAACAATTAACAATTAACAATTAACA 2733
Qy      933 ThrValAlaAaPValAaPValAaPValAaPValAaPValAaPValAaPValAaPValAaPVal 952
Db      2734 ACCATTAAAGCGCAAGGCAACAGACTGCACCTTTACAGAACACAGATCACTGGGTC 2793
Qy      953 ThrGlnThrSerThrHISThrSerThrSerThrThrThrThrThrThrThrThrThrThrThr 972
Db      2794 ACAAAAGCGTGCACCATCACTACGACATCCACAACACATTAAGTACAGTCAAGATA 2853
Qy      973 ThrLeuThrSerThrAaPValAaPValAaPValAaPValAaPValAaPValAaPValAaPVal 992
Db      2854 ACCTTGAACCTGCACAGCGGTGTAAGCTTGAAGCTGTAAGCTGTAAGCTGTAAGCTGTAAG 2913
Qy      993 AlaGlyAaPValAaPValAaPValAaPValAaPValAaPValAaPValAaPValAaPValAaPVal 1012
Db      2914 GCAAGAGACGTAAGAAAGGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG 2973
Qy      1013 ValLeuValAaPValAaPValAaPValAaPValAaPValAaPValAaPValAaPValAaPVal 1032
Db      2974 GTGTTATTAGCAATGACGATTTCAATCAATGAT 3006

```

RESULT 8
US-07-781-034-1
Sequence 1, Application US/07781034

```

; Patent No. 5442050
; GENERAL INFORMATION:
; APPLICANT: Fishman, Jay A.
; TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
; TITLE OF INVENTION: Rat- and Human-Derived Pneumocystis Carinii
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/781,034
; FILING DATE: 19911018
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MGH91-02A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2814 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1617..2813
; US-07-781-034-1
;
; Alignment Scores:
; Pred. No.: 1 666-88 Length: 2814
; Score: 913.00 Matches: 211
; Percent Similarity: 53.99% Conservative: 107
; Best Local Similarity: 35.82% Mismatches: 223
; Query Match: 16.87% Indels: 48
; DB: 1 Gaps: 13
;
; US-10-654-416-14 (1-1023) x US-07-781-034-1 (1-2814)
;
; Qy      5 ValLysArgGlnAla-----AlaGlyThrGlnAsnSerLysAaPValAaPValAaPVal 21
; Db      1131 GTTAAGAGCGCAAGACAAAGTATGAACATGACACAGATGAGATTAAAGAGAAACACTT 1190
;
; Qy      22 LeuAlaLeuLysLysGluAaPValAaPValAaPValAaPValAaPValAaPValAaPVal 41
; Db      1191 TTGGCTTCACTTGTGAAGCAAAACATGATGATGAGATGAATGAATGAATGAATGAATGA 1250
;
; Qy      42 LysTyrCysGlnGluLeuThrGluAlaLysLeuAaPValAaPValAaPValAaPValAaPVal 61
; Db      1251 GAATATTGTAAGAGGTTGAAGAAAGCAAGATGAGAAATTTCC--AGTGAATGAGAAAGTT 1307
;
; Qy      62 LysGlyPheCysGluAaPValAaPValAaPValAaPValAaPValAaPValAaPValAaPVal 81
; Db      1308 AAAGACTTGTGTATGATTAATAAAGCAAGCAAAATGCAAAAGAACTGAAGAAAAAGTT 1367
;
; Qy      82 GluLysLysCysThrThrThrLysGlyLysLeuLysGluAlaLysLysLysLysLysLysLys 101
; Db      1368 AAAGATGAATTTGGAACTTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1418

```



```

Oy 11ellethrapplyaspcyelyegluabngluginCybleuPheleugluVal 121
Db 1419 ATGAAAGATGAAAGATTGTGTAAGAAATGATATCTTTAGAGAAACA 1478
Oy 122 CysSerlysgluLeuLeuAspAspCysAsnThrLeuArgAsnlyCysTyrglnlysls 141
Db 1479 GACCAGAAATAGCTTTAAGAGAACTGTGTCAAGTTGAGGAAAGATGTTCAGAAATGAAAG 1538
Oy 142 ArgAspLyValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuabnglyser 161
Db 1539 CGTAAAAAGGTGGCAGAGAGCTCTTTGAGGGGCTCGAGAGGAGATCTTAAGATGAA 1598
Oy 162 ValIleCysGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 181
Db 1599 GCTAAATGTAAGAAAGAAAGTGAAGAACTGTTCCTCAAGTTAAGCGGAAAGTACAG 1658
Oy 182 LeuThrAsnLeuCysLeuAsnGlnlysgluthrCysLyAsnIleLeuIleGluLyAsp 201
Db 1659 CTGATGTTTCTGCTGCTGATTCGAGTGAACGTCTAAAGCGCTGAAACAAATCAGAA 1718
Oy 202 LyblyCysGlyThrLeuLeuThrAspValSerAlaAlaLeuGlySerPheLyblyGlu 221
Db 1719 GAAAGTTGCTGCTCTTAAAGAAAGCTTAA-----GATGCGCAATTAAAGGAA 1769
Oy 222 ThrCysLeuGluLeuLeuGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 241
Db 1770 AAATGTCATGAAAGCTGAGAAATGCTATTTTCAAGAAAGCGTGAAGCTGAA----- 1823
Oy 242 IleIleLeuCysIleGluLeuGlyLyblyCysGlnGluGlnAsnIleAlaTyrePro 261
Db 1824 ---ACAAAGTGTGATGAGAGATGTAAGCAATGCAAGAAAGAAAGATTCACATATAAGCG 1880
Oy 262 ProGlyProAspPheAspProThrArgProGluAlaThrIleAlaGluAspIleGlyLeu 281
Db 1881 CCGGATCTGATTTAGTCTCTGCAAGCGGAGCGGCTGCTGTGAGAAAGTATGGCTTG 1940
Oy 282 GluGluPheTyrlLyblyValGluGluAspGlyValPheIleGlyAsnHleu---- 300
Db 1941 GATGATGTGATTAAGAAAGCTGAAAGAGAAATTAATTATTTGAAAGATCGAGATGAT 2000
Oy 301 -----ArgAspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLy 317
Db 2001 CTACCAAGAAAGTCAAGTAAACAATTTCTGCAAGATCTCTGCTGATGAGCAAGAT 2060
Oy 318 LyblyAspAspLyblyGluLyblyCysGluGluAlaLeuGlnlySerCysLybAsnProHs 337
Db 2061 GAGATGTATGCGAGGAGAAAGTCCGTTAAGCGTTAAGAAATGTAAGCTTCTAAGTAT 2120
Oy 338 GluHsGluAlaLeuGluSerLeuCysLyblyAsnGlyLeuSerAsnAspGlyThrLy 357
Db 2121 TTGAATCTGATTTGATGAGATTAAGCA-----GATGCTGATTAAGAAAT 2168
Oy 358 LyblyCysGluGluLeuGlnAsnAspIleAsnlyThrCysLyblyPheThrSerLyVal 377
Db 2169 AAATGCAAAAAA---AGCTGATGTAAAGAAAGATG----- 2204
Oy 378 ThrAsnAsnArgLeuPheAspProThrLybly----- 388
Db 2205 ACMAAACTCAAGTTAATCTTATGTGAAGAGGTGTCTACGAGTTTAAAGAGATAA 2264
Oy 389 AsnAsnGluIleValGlyTrpGluGlyLeuProThrPheLeuSerAsnGluAspCysAla 408
Db 2265 AAATCAACTCTTTATCGTGGGACAGCTTCCAAACATTAATTTAAGAGGAGAGTGTCA 2324
Oy 409 LyblyLeuGluSerTyCysPheTyPheGluLyblyCysProAspGlyGlu-----Asn 426
Db 2325 GAACTTGAGTGGAGATGTTCTATTTAAGAAATGCGTGAAGAAATGAAGATGTGTGA 2384
Oy 427 AlaCysLyblyAsnIleArgAlaThrCysTyrlLyblyAspGlyLeuAspAlaArgAlaAsnly 446
Db 2385 GCGTGTCAAAATCTACGATCAAGCGGTATTAAGAGGAGCAAGACAGATGTTGAATAG 2444
Oy 447 ValIleGluGlnGluAsnMetArgGlyMetLeu---HisGlySerAsnlySerTrpLeuGlu 465

```

```

Db 2445 TTCTTCAAAAAGAAATGMAAGGAAAGCTGCTATGTAAGATTTTATAGGATCTTAA 2504
Oy 466 LybPheGlnGlnGluLeuValCysGluLyblyLeuLyblyGluAsnlyGlySer 485
Db 2505 GATTGTAAAAAATATGTGTGTAAGAACTGTACAAACCTTAAGAAAGATTAAGAA----- 2558
Oy 486 PheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLyblyAlaAlaArgLeuLeu 505
Db 2559 -----TACCTTTCAAAATGTCTTATCTTAAGAAAGCTATGTGTTATGGCTT 2603
Oy 506 ThrAspLeuArgMetLyThrIlePheLeuArgGlnGlnLeuAspGlnLyblyAsp 525
Db 2604 TCAATATATATTTTCTCCAAATCCAAAGAGTTAAGTTCGCTTTAGATGATCAGAGAT 2663
Oy 526 PheProThrAspLyblyAsnCysLyblyGluLeuGlyArgLyblyCysGlnAspLeuGlyAsp 545
Db 2664 TTTCATTGAAAGAAAGATGTGCTGAATGGGAGAGAGTGTGATCAACTTATAGTAT 2723
Oy 546 SerLyblyGluLeuTrpProCysHisThrLeuGluGlnGlnCysAsnArgLeuGlyThr 565
Db 2724 TCATTTATTTGAATTTAAGAAAGTGTATTAACATTTGAAGAAAGCTGTGATTTTGAAGCTT 2783
Oy 566 ThrGluLeuLeuLyblyGlnValLeuLeu 574
Db 2784 ACAGAAAGATTTAGAAAGATTTTATTTTA 2810

RESULT 9
PCT-US92-08328-1
Sequence 1, Application PC/TUS9208328
GENERAL INFORMATION:
APPLICANT: Fishman, Jay A.
TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
NUMBER OF INVENTION: Rat- and Human-Derived Pneumocystis Carinii
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Miltia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/08328
FILING DATE: 19920930
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/781,034
FILING DATE: 18-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,166
FILING DATE: 30-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MGH91-02AA PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
TELEX: 951794
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2814 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO

```



```

; ANTI-SENSE: NO
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1617..2813
;
PCT-US92-08328-1

```

Alignment Scores:	
Pred. No.:	1,66e-88
Score:	913.00
Percent Similarity:	53.9%
Best Local Similarity:	35.82%
Query Match:	16.87%
DB:	5
Length:	281
Matches:	217
Conservative:	101
Mismatches:	223
Indels:	48
Gaps:	13

US-10-654-416-14 (1-1023) X PCT-US92-08328-1 (1-2814)

Qy	5	VallyvArpGlnIaIa-----AlaglyThrGlnAmserIleAerGlnGlnHleVal	21
Db	1131	GTTAAGAGGCAAGCAAGTAAGTACAGTACACCAAGATGAGATTAAAGAGAAACACTT	1190
Qy	22	LeuAlaLeuIleLeuLySGluAerGlyLeuSerGlnGlnIuCyLyLyLyLeuLyLy	41
Db	1191	TTGGCTTTCATTGTGGAAGGACAAACATGATGATGAATGAATGCAAAAAAAGGCTCGAG	1250
Qy	42	LySerLyCySGlnGlnIuLeuThrGlnAlaLyLeuAmIleGlnGlnValHleAArgLyLeu	61
Db	1251	GAATATTTGTAAGAGTTGAAGAAAGACAGATGAGATTTTC---AGTGTGAATGAGAAAGTT	1307
Qy	62	LySerLyPheCySGluAerGlyLyLeAlaAerThrLyCyLySGluLeuLyAlaAerIle	81
Db	1308	AAAGGACCTTTGTGTGATGATAAAAACGAGACGAAATATCAAGAACTGAATAAAAAAGTT	1367
Qy	82	GluLyLyLyCySerThrThrIleLySerLyLyLeuLySGluAlaIleLyLyLyIleGln	101
Db	1368	AAATATGAATTTGGGAACCTTTTGATATCGAGATCTTGACATCGATGATATGAT-----GAC	1418
Qy	102	IleIleThrAerLyAerAerCyLyAerGlnAmGlnGlnGlnIuCyLeuPheLeuGlnGlyAl	121
Db	1419	ATTAAGATGAGAGAGATTGTATTAATAACATAGAAATAATGATATCTTTTAGAGAGACA	1478
Qy	122	CySerLySGluLeuLyAerAerCySerAmThrLeuAerGlnLyCySerLyGlnLyLyLy	141
Db	1479	GACCAATATAGCTTTAAGAGAACTGTGTCAAGTTGAGAGAAAGATGTTACGAATTGAG	1538
Qy	142	ArgAerLySerValAlaGlnGlnValLeuLeuArgAlaLeuAerSerAerLeuAmGlySer	161
Db	1539	CGTAAAAAGCGGAGAGAGACTCTTTTGAAGGCGCTCGAGGGGAGCTTAAAGATGAA	1598
Qy	162	ValIleCySGlnLyLyLeuLySGluIleCySerProValIleGlyArgGlnSerAerGln	181
Db	1599	GCTAAATGTAAAGAAAGATGATAAAGTGTGGCCCAAGTTTAAGCCGAAAGATGACAG	1658
Qy	182	LeuThrAerLeuCyLeuAmGlnLySGluThrCyLyAerAmIleLeuIleGlnLyAer	201
Db	1659	CTGATGTTTTTCTGCTCTGATTCGGATGGAACGTGTAAAGCGCTGAACAAATACGAA	1718
Qy	202	LyLyLyCySGlnLyThrLeuLyThrAerValSerAlaAlaLeuGlySerPheLyLySGln	221
Db	1719	GAATTTTGCTGCTTTTAAAGAAAGGCTTAA-----GATGCGAATTAAAGGA	1769
Qy	222	ThrCyLeuGlnLeuLeuGlnGlnIuCySerPheThrLyIleGlyAmCyGlyLyAerAerAer	241
Db	1770	AAATGTCTAGAAAGACTTGGAATGTGATTTTACAAAGAAAGGTGATCGAA-----	1823
Qy	242	IleIleLyCyGlyIleGlnLeuGlyLyLyCySGlnGlnGlnAmIleAlaTyrmecPro	261
Db	1824	---ACAAAGTGTGATGAGATATGAAAGCAATGCAAGAAAGAAAGATTCACTTTAAACG	1880
Qy	262	ProGlyProAerPheAerProThrAerProGlnAlaThrIleAlaGlnAerGlyLyLeu	281
Db	1881	CCGGAATGTGATTTTAAGTCTCTGTCAAGCCGAAGGCTGCTTGTGAAGAAATTCGGTTG	1940
Qy	282	GlnGlnPheThrLyLyLyValGlnGlnAerGlyAlaPheIleGlyLyAerAmIleLeu---	300

[illegible]

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 STREET: Two Milltia Drive
 CITY: Lexington
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02173

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/781,034
 FILING DATE: 19911018
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Granahan, Patricia
 REGISTRATION NUMBER: 32,227
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 861-6240
 TELEFAX: (617) 861-9540
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1448 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 US-07-781-034-3

Alignment Scores:

Prod. No.: 1,986-50 Length: 1448
 Score: 559.00 Matches: 136
 Percent Similarity: 55.87% Conservative: 59
 Best Local Similarity: 38.97% Mismatches: 117
 Query Match: 10.33% Indels: 38
 DB: 1 Gaps: 8

US-10-654-416-14 (1-1023) * US-07-781-034-3 (1-1448)

QY 1 MetAlaTgAlaValLysArgGlnAlaAglYThrGlnAenSerIle-Aap----- 17
 DB 388 ATGGACCGCGGTTAAAGGCAACCAAGTAAAGTAAAGCAAAATGATGAGATTAA 447
 QY 18 -GlUcHlStValLeuAlaLeuLleLeuLysGluAapGlyLeuSerGluGlnGluValY 37
 DB 448 GGAGGAAACCTTTGGCTTTCATTGTGAGACAAACATGATGATGAGAAATGCA 507
 QY 37 slyeLysLeuLysTyRcYsGlnGluLeuThrGluAlaLysLeuAenIleGluGlnVa 57
 DB 508 AAAAAGCTTCAGAAATTTGTAAAGATTGAAGAAACAGATGGAATTTTC---AGCT 564
 QY 57 lHieArgLysLeuLysGlyPheCysGluAapGlyLysAlaAapThrLysCysLysGluLe 77
 DB 565 GAATGAGAAAGTAAAGACTTTGTGATGATTAAGAAACGAGCAAAATGCAGAAAGACT 624
 QY 77 uLyAlaAenIleGluLysCysThrThrLysGlyLeuLysLeuLysGluAlaLleLys 97
 DB 625 GAAAAAAGCTGGGATGGAATTTGATACGATCTTGAGCAATCGGTAGA 684
 QY 97 slyeLysIleGlnIleLeuThrAapLysAapCysLysGluAenGluGlnGlnCysLeuPh 117
 DB 685 T-----GACATGGAAGATGAAGAGCTTTGTAAGAAACATGAAGAAATGATATCT 735
 QY 117 eLeuGluGluValCysSerLysGluLeuLysAapAapCysAenThrLysAenLysCy 137
 DB 736 TTTAGAGAGAGAGACCAAAATAGCTTAAAGAGAACTGTGCAAGTTGAGGAGAGATG 795
 QY 137 eTyGlnLysLysAapLysValAlaGluGluValLeuLeuAgaAlaLeuArgSerAs 157

DB 796 TTACGAATTGAGCGTAAAAAGGTGCGAGAGAGCTCTTTTGAAGGCGCTCGAAGAGA 855
 QY 157 pLeuAenGlySerValIleCysGluLysLysLeuLysGluLleCysAProValMetGlyAr 177
 DB 856 AGCTAAAGAAAGATTAAATGTAAGACAGAGATGAAAGAGTTGGCCAGAGTTAAAGCG 915
 QY 177 gGluSerAapGluLeuThrAenLeuCysLeuAenGluLysGluLthrCysLysAenIleLe 197
 DB 916 AGAAAGGACGAATGATGTTTGTGCTTGAATTCGATGCAACGCTGCGCCCTGAAA 975
 QY 197 uIleGluLysAapLys-----LysCysGlyThrLeuLysTh 209
 DB 976 AAAAAATCAGAAAGAGTTTGCACAGCTTTTAAAGAAAGCTTAAAGATGCGCAATTA----- 1030
 QY 209 rAapValSerAlaLeuGlySerPheLysLysGlyThrCysLeuGluLeuLeuGluGlu 229
 DB 1031 -----AAGCAAAATGTCATGAAAGACTTGAGAA 1059
 QY 229 nCyRTrPheTyrlleGlyAenCysGlyAapAapLlelleLysCysIleGluLeuGlu 249
 DB 1060 ATGTCATTTTAAACGAGAGAGCTGT-----GATAAACAAAAATGTGATGAGATTA 1110
 QY 249 yGlyLysCysGlnGluGlnAenIleAlaTyRMetProProGlyProAapPheAapProTh 269
 DB 1111 GGATCAATGCGAGAAAAAAGAAATACATATA--GCCCGCAATCTGATTTCTAGTCTGT 1169
 QY 269 rArgProGluAlaThrIleAlaGluAapIleGlyLeuGluGluLubThTyRlysValGl 289
 DB 1170 CAAGCCGAGAGCGTGTGTTGAGAGATTAATGGTGTGATGATGATGATTAAGAGCTGA 1229
 QY 289 uGluAapGlyValPheIleGlyLysAenHleLeu-----ArgAapAlaThrAl 305
 DB 1230 AAAAGAAAGAAATTTATTTGAAAAATCAGAGGTGATCTACCAAGAGAAAGTCAGATCAAA 1289
 QY 305 aLeuLeu-----AlaLeuLeuIleGlnAapSerSerLeuLysLysAapAapLysGl 323
 DB 1290 ATTCTGCAAGATCTCTTCTAGTCTTGACGACAGAGATGAGAAATGATTAAGATGCAAGGAA 1349
 QY 323 uLysCysGlnGluAlaLeuGlnLys 331
 DB 1350 GAATGCGAAAAAGCGTTAAAAAAA 1374

RESULT 11
 PCT-US92-08328-3
 Sequence 3, Application PC/TUS9208328
 GENERAL INFORMATION:
 APPLICANT: Fishman, Jay A.
 TITLE OF INVENTION: Molecular Cloning of Antigens Shared By
 TITLE OF INVENTION: Rat- and Human-Derived Pneumocystis Carinii
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 STREET: Two Milltia Drive
 CITY: Lexington
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02173
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/08328
 FILING DATE: 19920930
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/781,034
 FILING DATE: 18-OCT-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/768,166
 FILING DATE: 30-SEP-1991
 ATTORNEY/AGENT INFORMATION:

```

NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MGH91-02AA PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
TELEX: 951794
INFORMATION FOR SEQ. ID NO.: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1448 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US92-08328-3

Alignment Scores:
Pred. No.: 1.98e-50
Score: 559.00
Percent Similarity: 55.87%
Best Local Similarity: 38.97%
Query Match: 10.33%
DB: 5

US-10-654-416-14 (1-1023) x PCT-US92-08328-3 (1-1448)
Length: 1448
Matches: 136
Conservative: 59
Mismatches: 117
Indels: 38
Gaps: 8

```

[illegible]

```

Db      1031  -----AGGAAAAATGTCATGAAAGACTTGGAAA 1059
Qy      229  nCyStyrPheTyrIleGlyAsnCysGlyAspAspAspIleIleLysCysIleGluLeuG1 249
      :|||:|||||
Db      1060  ATGTCATTTTACGGAGAGCGTGT-----GATTAACAATAAGTGTAGAGATTA 1110
Qy      249  yGlyLysCysGlnGluGlnAsnIleAlaTyrMetProProGlyProAspPheAspProth 269
      :|||:|||||
Db      1111  GGATCATTGCGAGAAAAAAGAAATCATCATATTA -CGCCAGAAATCTGATTCTACTGCTGT 1169
Qy      269  rArgProGluIleThrIleAlaGluAspIleGlyLeuGlnGluPheTyrIleLysValG1 289
      :|||:|||||
Db      1170  CAAGCCGAAGACGCTGTTGTGAAGAAATATTTGGCTTGATGATGTGTATTAAGAAGCTGA 1229
Qy      289  uGluAspGlyValPheIleGlyLysAsnHisLeu-----ArgAspAlaThrAl 305
      :|||:|||||
Db      1230  AAAAAGAAAGAAATTTATTTTGGAATAATACGAGGTGATCTACCAAGGAAGCTCAGGTACAA 1289
Qy      305  aLeuLeu-----AlaLeuLeuIleGlnAspSerLeuLysIleLysAspAspLysG1 323
      :|||:|||||
Db      1290  ATTTTCTCAAGANTCTCTTGCTAGAGCTTGGACGACAGATGAGAATGATTAAGATGACAGGGA 1349
Qy      323  uLysCysGlnGluIleLeuGlnLys 331
      :|||:|||||
Db      1350  GAAATGCGAAAAAGCGTTAAAAAAA 1374

```

```

RESULT 12
US-09-762-724-15
; Sequence 15, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii
; PILE REFERENCE: 4739-58054
CURRENT APPLICATION NUMBER: US/09/762,724
CURRENT FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: PCT/US99/18750
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: US 60/096,805
PRIOR FILING DATE: 1998-08-17
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 249
TYPE: DNA
ORGANISM: Pneumocystis carinii sp. f. hominis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(249)
US-09-762-724-15

Alignment Scores:
Pred. No.:          5.38e-34      Length:           249
Score:              396.00       Matches:            73
Percent Similarity: 98.80%     Conservative:      9
Best Local Similarity: 87.95%   Mismatches:        1
Query Match:         7.32%      Indels:             0
DB:                  4          Gaps:               0

US-10-654-416-14 (1-1023) x US-09-762-724-15 (1-249)
Qy      GluCyAlvSeerLeuglnTrHrThAspThrTriValThrglnThSerThRHisThSer    960
Db      1 GAAGGCAAAATCTCGACGACGACGACACTGGGTCCAAAGAAGCTGCACCCTACTTACG    60
Qy      ThzSerThrlleThSerThrlleThSerIyslleThrDeuthSerThArgArgCys     980
Db      61 ACTCTACGACTAGCTCACGTCACATCGAATAACACTCACTCACTCAAGCGGCGTGTT    120
Qy      LysProThrlYscYsrThrlTrgLYgluGluAspAspAlaGlyAspValLysProSerGlu   1000

```


Dh	4087	GAGACTAAGCCAAAGACACTCTG69CCATGCGCTGAGCTGACGACCCGACATGACTGAGAC	414.6
Qy	515	PheLeuArlrgInglInLeuAepGlnLyaArgAsp-----	525
Dh	4147	CTGCTGCGGGAAACAGTATGAGAGAGGACGAGAAAGCCAAGCTGAGCTGACAGGGGAATG	420.6
Qy	526	-----PheProThrAsp-----	529
Dh	4207	TCCAAGCCCAACAGTAGAGCTTGCCCAAGTGGAGACCAAGTACGAGACGCCAATCCAG	426.6
Qy	530	-----LysAsnCysLysGluLeuGluIArgLysCysGlnAepLeuGly	543
Dh	4267	CGCACAGAGAGAGCTGGAGAGAGCCCAAGAAAGAAAGCTAAGCCAGCGCTGTCAGAGATGCAAGAA	432.6
Qy	544	GluAepSerLysGluLLeThrTrpProCysHleThrLeuGluGlnGlnCysAsnArlGlu	563
Dh	4327	GAACATGTAGAACGCTGAATTCCAAAATGTCTCTCTTGTGAAGAAG-----	437.1
Qy	564	GlyThrThrGluLLeuLysGlnValLeuLeuAepGlnHleLysAepThrLeuLysAep	583
Dh	4372	-----ACAAAGCAGAGGCTTACAGAAATGAATGAATGAGAGAGACTCATGATTTGAT	441.6
Qy	584	GlnGlu-----SerCysValLysTrpLeuLysGluLysCysAsn-----	596
Dh	4417	GTCGAACGATCTATATGCTGCTGCTGACTAGCTCTCATTAAGAAAGAAACTTTGACAAAG	447.6
Qy	597	-----LysTrpSerArlArgGlyAAspArlArgPheSerPheValCysValPheGln	613
Dh	4477	GTTCTGCAGAAATGAAACAGAG-----TATGAG	450.6
Qy	614	AsnArlArlCysGluLeuMetValLysAepValLysArlArg-----CysGlnVal	630
Dh	4507	GAAACTCAGGCTGAACCTTGAGAGCCCTCCCAAGAGAGTGCGCTTCTTCAGACATGAGCTG	456.6
Qy	631	PheLysLysAenLLeLysAlaSerTrpLleLeuGluAAsnAAsnThrAAsnLys	650
Dh	4567	TTCAAG-----GTGAAGATGCTTAC-----GAGAAATCCCTGATAT	460.5
Qy	651	LleThrThrLeuGluArlArgAsnCysAepSerTrpHleThrTrpCysAAsnArlPheSerPro	670
Dh	4606	CTTGAAACTCTAAAGCGAGAG-----AATTAAGACTTACAAACAG	464.4
Qy	671	AsnCysProGluLLeuThrLysGlu-----AsnSerCysThr	682
Dh	4645	GAGATTTCTGACCTGACGACAGCAAAATTCAGAGGGTGAAAGCATATTCATGAACTGAG	470.4
Qy	683	LysLLeLysLys-----	686
Dh	4705	AAATGTAAGAAACCACTTGATCATGAGAAAGCTGAATCAACAGACTTCCCTTAGAGAAAGCA	476.4
Qy	687	-----HisArlGluProPheTrpLysArlGlyAlaLeuGlu-----AspAla	700
Dh	4765	GAGCATCTCTTGAAGCATGAAGAAAGGCAAAATTTCTGCCATTCAACTTGACACTAAATCAG	482.4
Qy	701	LeuLysValGluLeuGlnGlyLysLeuThrArlLysSerLysCysGluProAlaLeuLys	720
Dh	4825	GTGAATCTGAGATTGACCGCAAAATTTCTGTAATAAAGATGAAGAACTGATCAGCTAAAG	488.4
Qy	721	Arg-----	721
Dh	4885	AGGAACCATCTCAGAGTTGTGAGACTCAATGCAGAGTACATCGATCTGAGATCAGAGAC	494.4
Qy	722	-----TyrCysThrValAlaGlyAAsnValAsnAAsnAlaSerLle	734
Dh	4945	AGAAATGATGCTCTGAGAGTCAAGAAAGAAATGAGAGGAGATCTTAAATGAATGGAATTC	500.4
Qy	735	SerGlyLeuCysLysAlaAsnThrLys-----	743
Dh	5005	CAG---CTGAACCAATGCCAACCGCAGCGCTGCTGAGCAGCTAAGGAATCTTAGAAACACA	506.1
Qy	744	-----AspAsnSerGlyLysSerArlAepGluArl	752
Dh	5062	CAAGGAATCTGAAGAGACATCAGCTTACATTTGATGTATGCTATTCAGAGGCGCAAGATATAC	512.1

```

Qy      753 AlaArglyseLluCyvsgVlluvlBleValValglvgLuvcylvslalaleu    772
Db      5122 CTTAAGGAAACACTGGCA-----ATGGTTCAGCGCAGAGCTTAACCTGATG    5166
Qy      773 ProThrGlueugLygnPrcoAlaAlaaPrlLeuLyvsApRyTrlvThrTygIuGlU    792
Db      5167 CAGGCTGAATTGAAGAAGCTCAGGAGCATCCCTGGAACGCACT-----GAGAGA    5214
Qy      793 LeuLylyeArAlagIuGlUAlameArAsnLysSerSeuValLeuSerlle---    811
Db      5215 GGCAGAGAAAATGGCAGAGCAAAGACTTGGAATGCCGTMAACGTGCAACTTCGAC    5274
Qy      812 -----LylvYasngIUsenAenValSerlys--    820
Db      5275 ACTCAGAACCCAGACTGATCAACACCAGAAAGAGCTGAAACAGCATTTCCCAATC    5334
Qy      821 -----SerAnSerlysaenlysaPryAsn    829
Db      5335 CAGGAGAGATGGAGCACATCGTCCAGAAAGCCCGCAATGCAGAGGAGGACCAAG    5394
Qy      830 AlavalserhenglyLeu-----GlnaPrrThrLys    840
Db      5395 GCCATCACTGATGCTGCCATGATGCTAGAGAGCTGAGAGAGAAACAGACACAGCGCC    5454
Qy      841 HisvalylleuAargrgly-----VallysaPryVal-----    852
Db      5455 CACCTGAGCGGATGAAAGAAACAATGAGACACAGCCTGAAAGATCTGCAGACTCCGCTG    5514
Qy      853 -----SerValThrGlueugLualA    859
Db      5515 GATGAGGCTGAGCAGCATGCGCGCTGAAGAGGTGGGAAGAACAGATCCAGAAACTGGAAGCC    5574
Qy      860 LylaIaherpleuAlaAlaglValPhcglYArgYrValser-----    874
Db      5575 AGGCTGAGAGAGCTTGAAGTAGGTGGAAGTGAACAGAAACAATGTTGAGGCTGTC    5634
Qy      875 -----LeuLygluArgCyvaenlyvleuGuJusePsr    885
Db      5635 AAGGCTTTCGCAAAACATGAGAGAAAGAGTGAAGAACTCATCAACAGACTGAGAGAGAC    5694
Qy      886 CyvalArglyleugLuarCyvlysaPryleugLugalValCyvlyelysilleenlysaI    905
Db      5695 CGCAAGAAATATTCACAGGCTGCAGAGCTTGATGACAATAATGGCAAAAGTCAAAAGCT    5754
Qy      905 acYvalArgsn---LeuLyProlenugLualLysPronIsgIuThrValThrgIusePth    924
Db      5755 TACAAGAGACAAGGTGAAGAGGCTGAGAAACAATCCAATGTCAACTTGGCAAGTTCGC    5814
Qy      924 rThrThrThrThrThr-----ThrThrThrValAlaAPrrProlysaI    939
Db      5815 AAGCTCAGACACAGCTGAGAGAGGCGGAGAGAACGGGCTGACATGTGTAGTCCCAAGTC    5874
Qy      939 arHr 940
Db      5875 AACCA 5878

RESULT_15
US-09-750-590A-3
; Sequence 3, Application US/09750590A
; Patent No. 6780987
; GENERAL INFORMATION:
; APPLICANT: Herman, Ira
; TITLE OF INVENTION: Beta-Car73 CONTROL OF NORMAL AND
; ABNORMAL CELL MIGRATION
; FILE REFERENCE: TUI-001CP
; CURRENT APPLICATION NUMBER: US/09/750.590A
; PRIORITY FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: 60/170,182
; PRIOR FILING DATE: 1999-12-10
; PRIOR APPLICATION NUMBER: 09/733,818
; PRIOR FILING DATE: 2000-12-08

```

RESULT 15
US-09-750-590A-3
Sequence 3, Application US/09750590A
Patent No. 6780987
GENERAL INFORMATION:
APPLICANT: Welch, Alice
APPLICANT: Herman, Ira
TITLE OF INVENTION: BETA-CAP73 CONTROL OF
TITLE OF INVENTION: ABNORMAL CELL MIGRATI
FILE REFERENCE: TUI-001CP
CURRENT APPLICATION NUMBER: US/09/750,590A
CURRENT FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: 60/110,182
PRIOR FILING DATE: 1999-12-10
PRIOR APPLICATION NUMBER: 09/753,818
PRIOR FILING DATE: 2000-12-08

```

; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 4206
; TYPE: DNA
; ORGANISM: Bos taurus
US-09-750-590A-3

Alignment Scores:
Pred. No.: 2,936-17 Length: 4206
Score: 262.00 Matches: 226
Percent Similarity: 34.94% Conservative: 151
Best Local Similarity: 20.95% Mismatches: 380
Query Match: 4.84% Indels: 323
DB: Gaps: 46

US-10-654-416-14 (1-1023) x US-09-750-590A-3 (1-4206)

OY 5 VallysaArgGlnAlaAlaGlyThrGln-----AenSerIleaspGluGlu 19
DB 1606 GTGAAATACGAAGAGTCCGCGCAGAAAGTGGGAAATTGAGAAACCAATCAACAAAT 1665
OY 20 HisValIleuAlaIleuIleuIleuIleuSerGluIleuSerGluGluGluCysIleuIleu 39
DB 1666 GAATGTTAGTTGAAGAGATTAAAGAGATGAGGGCAAGCTGATGAAAGAGAAATTAAGCGA 1725
OY 40 LeuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 59
DB 1726 CTGCAGAG-----GAGTTGAGCATGTCTGAATCTGAGCCGAGAGAGAGAGAGAGAG 1776
OY 60 IysIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 79
DB 1777 AAGCTCACT-----GAGATGAGAGGCCAGTTTAAAGACTGTGCACC 1818
OY 80 AenIle-----GluIleIleIleIleIleIleIleIleIleIleIleIle 90
DB 1819 AAGCTGGCCCTTCTATTCACAGAGAAATTTGAAACATGAAAGCTTGTATCAAT 1878
OY 91 IysIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 110
DB 1879 GAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1938
OY 111 AenGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 130
DB 1939 AATGAA-----ACTAGCATTAAAGAGAACTT 1968
OY 131 AenThrIleuArgAenIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 150
DB 1969 GAGAAATTTGAAGGCCAACTGGCTCAGACCTCAACCAAGAGAGAGAGAGAGAGAGAGAG 2028
OY 151 LeuArg-----AlaIleuArgSerAspIleuIleuIleuIleuIleuIleuIleuIleu 165
DB 2029 AGCAGATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088
OY 166 -----LysIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 180
DB 2089 AATCAGAGCTTACAAAGAGAAATCGAAAAGCTCTCGATTAATTAAGCTCTTACACAA 2148
OY 181 GluIleuThrAenIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 200
DB 2149 CAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2208
OY 201 AspIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 219
DB 2209 AAAAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2268
OY 220 -----LysGluThrCysIleuGluIleuIleuIleuIleuIleuIleuIleuIleu 237
DB 2269 AAATATACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2318
OY 238 GluAspAspAspIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 257
DB 2319 -----CAGTTTAAGTAAATATGTCAG----- 2339
```

```

OY 258 AlaTrpMetProPheGly-ProAspPheAspProThrArgProGluAlaThrIleAlaGly 277
DB 2340 -----CCGCTCGAAACTGTGTTCAATACCT-----CCCAAGAGACAGAGAAAGA 2384
OY 277 uAspIleGlyIleuGlu-----GluPheTyIleIleuValGluIleuAspGlyVal 293
DB 2385 AATGATGGCTCTGAAATCCAAATATCACTGAACTTAAAGAGCGAGCTGTGAA----- 2436
OY 293 I PheIleGlyIleuAsnIleuAsnIleuAsnIleuAsnIleuAsnIleuAsnIleuAsn 313
DB 2437 -----CTTAATTAATAAATGTGTGTAAGCAAGAGAAATATATTCACATCATGTCTGAAA 2492
OY 313 pSerSerIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 319
DB 2493 CAATGATTTGAAAAAGACATGAGTCATCATGATATGCCCCGTGAAAAACCATGAAAGAT 2552
OY 320 -----AspAspIleuGly 323
DB 2553 TAAAACTGCTTGATGACACATTGATTAATAAACAATAGAGAAATTAAGATGAGAGAA 2612
OY 323 uIleCysGluGluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 343
DB 2613 GAATGTGAAGATTAATATCAAGAAATTTGTGAAAAATAAAGATGAGAACGAAATATTA-- 2670
OY 343 uSerIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 354
DB 2671 -----AAAAAATCTGGAGAACATCTGAGAACCAACCAAGTAAAGCTGATCAT 2717
OY 355 -GlyThrIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 374
DB 2718 CAGCTTAAGAGAGATTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2769
OY 374 rSerIleuValThrAsnAsnArg-----LeuPheAspProThrIleuIleuIleuIleu 391
DB 2770 -----AAGTCCAGGACAAACAGCGCTGAATATCTGCTAAGTAAACCAACCAAGAGGA 2825
OY 391 uIleValGlyTrp-----GluGlyIleuProThrPheIleu 403
DB 2826 GATTGTACCCCTGATGAGAGATTCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
OY 403 rAsnGluAspCysAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 423
DB 2881 -----CAGAAATGATCAAGGCTAAATATGCTCCGATCATTCAGCTTGGAGAGGTGAG-- 2934
OY 423 pGlyIleuAsnAlaCysIleuAsnIleuArgAlaThrCysTyIleuIleuIleuIleuIleu 443
DB 2935 -----AGAAAATTTAAAGCCACT----- 2952
OY 443 gAlaAsnIleuValIleuGluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 463
DB 2953 -----GAGAAAGAACTTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3006
OY 463 pLeuGluIleuPheGluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 479
DB 3007 -----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3050
OY 479 sIleGluAsnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 499
DB 3051 GAAGAG-----ATCTCACTCTTCAG----- 3072
OY 499 aIleuAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 519
DB 3073 -----AAGGATCTTAAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 3110
OY 519 nIleuAspGluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 539
DB 3111 TTATGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3164
OY 539 sGlnAspIleuGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 559
DB 3165 AAAAGACTCTGTGAGAAATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3224
```


Qy 559 nCyAaenAArgLeuGlyThrThrGluLeu---LySGlnValLeuLeuAerGluH1sly 578
Db 3225 A---AATGCCAAGCAGACTCTTGAGATCTTGACAGCAAACTCTTTGAGAGAGCAGA 3281
Qy 578 aAerThrLeuLyAerGlnGlySerCyAValLyGlyLeuLySGlnCyAenLyStr 598
Db 3282 TGTTCGGGTGAGCAGGTGAGTCCCTGAAAAATCTCTT----- 3321
Qy 598 pSerArgArgGlyAerAerArgPheSerPheValCyAValPheGlnAenLyThrCyAGl 618
Db 3322 -----AGTGGTACATGCA 3335
Qy 618 uLeuAerValLyAerValLyAerArgCySGlnValPheLyAValAenLyLeys--Al 637
Db 3336 GACACTCAAGAGAACTGAAACTAAGCAGAGATGTTATGAGAAAGAGCAGCAGACAGT 3395
Qy 637 aSerTrtleileGluPheLeuGlnAenThrAenLyAValleThrThrLeuGluAVal 657
Db 3396 GACCCAACTGGGCGAGTGTGAGAAAT----- 3423
Qy 657 nCyAProSerTrpH1eThrTrpCyAaenAArgPheSerProAenCyAProGlyLeuThrLy 677
Db 3424 -----CA 3425
Qy 677 eGluAerSerCyAThrLyAValLyAValAerGluAerGluProPheTrpLyAArgLyAAla 697
Db 3426 GAAGAACTCTCTGAGCCCTGGGCTGAGCATTTTGACG-----GT 3464
Qy 697 uGluAerAlaLeuLyAValGluLeuGlnGlyLeuLeuThrAerLySerLyCySGluPr 717
Db 3465 TAAAGAAACATTTGAGAAAGAAATT----- 3489
Qy 717 oAlaLeuLyAerGlyTrpCyAThrValAlaGlyAenValAenAenLyAserLieserGlyLe 737
Db 3490 -----GGAAAT 3494
Qy 737 uCyAValAenAenThrLyAerAerAenSerGlyLySerAerGluAerAlaArgLyS--Gl 756
Db 3495 CATAAAGCTAGCTTGAGAGAA-----AAGAGAGAGAGAGCCAAAGCAAACTGA 3545
Qy 756 uLeuCySGlnLyAerValLySGlnValGluGlnGlnCyAlyAValAenProThrGluLe 776
Db 3546 AGAGGTCTCCAACTCCAGTCTGAGATTCAAGAACTAACAAGCGTTAAAAAATTAGA 3605
Qy 776 uGlyGlnProAlaAerLeuLyAerLyAerTrpLyThrTrpGluGluLeuLyAer 796
Db 3606 GACTCGGAGGTGCTGATTGTGCAAA--TATTAAGCA----- 3642
Qy 796 gAlaGluGlnAlaAerAenLySerSerLeuValLeuSerLeuLelyAerAenGlyLe 816
Db 3643 -----ACGAAAGCGATTG-----GAGAC 3662
Qy 816 rAenValSerLySerAenSerLyAenLyAerLyAerAlaValSerAenGlyLeuGl 836
Db 3663 ACACATTTCCGACTTAAGCAAAATTTGCAATCTGAATAGCAAGTATGAGCAAGTATG 3722
Qy 836 nAerThrThrLyAerValLyAerValLeuAerArgGlyValLyAerValSerValThrGl 856
Db 3723 TGAGAGGTTTGCATGCCAAAGAGAAAGCAAGCTGTCTGAAGAT-----GA 3770
Qy 856 uLeuGlnAlaLyAerAerAerLeuAlaGlnValPheGlyAArgLyValAerLeuLy 876
Db 3771 GAAGGAATGTGCTCATTTACACATAGAGCAAGAAATC-----AAAGATCAGCA 3818
Qy 876 eGluAerGlyAerLyAer-----LeuGlySerAerCyAArgLileLySGlnA 891
Db 3819 GGAACGATGTGCAAAATCTTAACAAACATCAGGAGCTACAGAGAAATACAGGAATC 3878
Qy 891 pCyAlyAerLyAerGluGlnValCyAlyAerLileAenLyAAlaCyAerAerLeuLyPr 911
Db 3879 TGCCAAACAAATCGAAAGCAAAAGATTAATGAATTAAGTGAATGCTCAATGATGTGAGAG 3938
Qy 911 oLeuGlu-----ValLyAerProH1eGluThrValThrGlySerThrThrThrThr 929

Db 3939 ATTAAACAGCCCTCAATGCGCTTTCCAGCTCACCCTATGGAAGTGGAGTCCAGCAA 3998
Qy 929 rThrThrThrValAlaAerProLyAAlaThrGluCyAlySerLeuGln--ThrThra 949
Db 3999 GAGCAGAGTACGTGATTGACAGCTTGACAGCAGCAGAGGTCAAGTCCCTGACAGCAGC-- 4054
Qy 949 sPrThrValThrGlnThr---SerThrH1eThrSerThrH1eThrSerThrH1 968
Db 4055 --ACGTGGCGATGCCGACAGACAGCAAGAGTAATTCATTTATCGACACACCT 4112
Qy 968 lThrSerLyAerLeuThrLeuThrSerThrAerArgCyAlyAerProThrLyAerCys 985
Db 4113 TCTTAGTCTGCACAGGCTCACATGATGAGATGAGATGAGGCGCGCTTACTGC 4165

Search completed: October 29, 2005, 15:19:52
Job time : 573 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: October 29, 2005, 09:48:47 ; Search time 1797 Seconds
(without alignments)
4701.696 Million cell updates/sec

Title: US-10-654-416-14
Perfect score: 5411
Sequence: 1 MARAVIRKQAAGTQNSIDEH.....MSGNVMRGVIVAVISFMI 1023

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9784742 seqs, 4129495052 residues

Total number of hits satisfying chosen parameters: 19569484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DB=ex1h
-O=/cgnt1/USPTO.pool/US10654416/runat_26102005_101025_21801/app_query.faeta_1.1159
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.tmpb -MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloum62
-TRANS=human40.cdd -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10654416@CGN_1.1159@runat_26102005_101025_21801
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA.*
1: /cgnt2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgnt2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgnt2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgnt2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgnt2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgnt2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgnt2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgnt2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgnt2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgnt2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgnt2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgnt2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgnt2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
14: /cgnt2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
15: /cgnt2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgnt2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
17: /cgnt2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
18: /cgnt2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgnt2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgnt2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgnt2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgnt2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq.*
23: /cgnt2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
24: /cgnt2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
25: /cgnt2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
26: /cgnt2_6/ptodata/1/pubpna/US11A_NEW_PUB.seq.*
27: /cgnt2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
28: /cgnt2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5411	100.0	3072	19 US-10-654-416-13	Sequence 13, App1
2	4398.5	81.3	3054	19 US-10-654-416-11	Sequence 11, App1
3	4352	80.4	3084	19 US-10-654-416-7	Sequence 7, App1
4	3827	70.7	3090	19 US-10-654-416-5	Sequence 5, App1
5	3401.5	62.9	3081	19 US-10-654-416-9	Sequence 9, App1
6	2752.5	51.1	3042	19 US-10-654-416-1	Sequence 1, App1
7	2404.5	44.4	3006	19 US-10-654-416-3	Sequence 3, App1
8	396	7.3	249	19 US-10-654-416-15	Sequence 15, App1
9	331.5	6.1	5784	18 US-10-282-122A-22545	Sequence 22545, A
10	312.5	5.8	5460	19 US-10-335-977-3230	Sequence 3219, Ap
11	311.5	5.8	5460	19 US-10-335-977-3230	Sequence 3220, Ap
12	308.5	5.7	5334	10 US-09-882-227-623	Sequence 623, App
13	274	5.1	5373	18 US-10-369-493-25273	Sequence 25273, A
14	274	5.1	5641	17 US-10-032-585-6646	Sequence 6646, Ap
15	273.5	5.1	3492	18 US-10-282-122A-15680	Sequence 15680, A
16	270.5	5.0	4087	18 US-10-369-493-26896	Sequence 26896, A
17	270	5.0	5886	21 US-10-798-037-3	Sequence 3, App1
18	270	5.0	6010	21 US-10-723-860-57	Sequence 57, App1
19	270	5.0	6010	21 US-10-723-860-105	Sequence 105, App
20	270	5.0	6010	21 US-10-723-860-4858	Sequence 4858, Ap
21	270	5.0	12801	21 US-10-798-037-5	Sequence 5, App1
22	269.5	5.0	4780	9 US-09-962-436-287	Sequence 287, App
23	269.5	5.0	4780	22 US-10-843-641A-2746	Sequence 2746, Ap
24	267.5	4.9	7985	26 US-11-097-143-9827	Sequence 9827, Ap
25	266	4.9	6965	21 US-10-335-053-225	Sequence 225, App
26	266	4.9	7694	17 US-10-096-534-34	Sequence 34, App1
27	266	4.9	7695	11 US-09-968-007A-462	Sequence 462, App1
28	266	4.9	7695	22 US-10-843-641A-6932	Sequence 6932, Ap
29	266	4.9	8063	10 US-09-814-353-21776	Sequence 21776, A
30	261	4.8	5835	10 US-09-927-597-1	Sequence 1, App1
31	261	4.8	5835	24 US-10-486-057-1	Sequence 1, App1
32	261	4.8	5937	10 US-09-927-597-3	Sequence 3, App1
33	261	4.8	5937	24 US-10-486-057-3	Sequence 3, App1
34	261	4.8	6861	15 US-10-171-311-161	Sequence 161, App
35	261	4.8	6861	18 US-10-341-434-102	Sequence 102, App
36	261	4.8	6861	24 US-10-923-035-18	Sequence 18, App
37	261	4.8	6900	15 US-10-171-311-163	Sequence 163, App
38	261	4.8	6900	24 US-10-764-425-13	Sequence 13, App1
39	260.5	4.8	5780	19 US-10-336-472-19	Sequence 19, App1
40	260.5	4.8	6773	9 US-09-864-864-136	Sequence 336, App
41	260	4.8	4428	18 US-10-395-027-445	Sequence 445, App
42	258.5	4.8	3570	19 US-10-336-472-15	Sequence 15, App1
43	258	4.8	11065	15 US-10-116-802-14	Sequence 14, App1
44	256	4.7	4643	9 US-09-864-864-266	Sequence 266, App
45	256	4.7	5956	24 US-10-783-271-100	Sequence 100, App

ALIGNMENTS

RESULT 1
US-10-654-416-13
; Sequence 13, Application US/10654416
; Publication No. US20040043409A1
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/10/654,416
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/09/762,724
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750

PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: US 60/096,805
PRIOR FILING DATE: 1998-08-17
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 3072
TYPE: DNA
ORGANISM: Pneumocystis carinii sp. f. hominis
FEATURE:
NAME/KEY: CDS
LOCATION: (1) ... (3072)
US-10-654-416-13

Alignment Scores:
Pred. No.: 0 Length: 3072
Score: 5411.00 Matches: 1023
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-10-654-416-14 (1-1023) x US-10-654-416-13 (1-3072)

QY	1	MetAlaArgAlaValLysArgGlnAlaAlaGlyThrGlnAsnSerTLeaSpGluLuhis	20
DB	1	ATGGGGCGGGCGGTACGCGGCGAGGAGGACAGAAATGACATTGATGAGGAAAT	60
QY	21	ValLeuAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu	40
DB	61	GTTTACGCTTTAATCTTAAGAGAAATGACCTAATGACAGGATGCAAAAAAATCTA	120
QY	41	LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys	60
DB	121	AAAAAATTTGGCAAGAAATTGACTGAGCAAAAATTAATATGAAACAGTACACAGAAA	180
QY	61	LeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAlaAsn	80
DB	181	CTTAAAGGTTTTGGCAAGATGAGAAAGACAGATACAAATGCAAAAGAACTGAAAGCCAT	240
QY	81	IleGluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLysLysIle	100
DB	241	ATTGAGAAATAATGATCTAATCAATCAAGGAAATCTTAAAGAAATTTAAAAAATTT	300
QY	101	GlnIleIleThrAspLysAspCysLysGluAsnGlnGlnGlnCysLeuPheLeuGluLys	120
DB	301	CAGATTATACGAGATACGATTCGAAAGAAATGAAACAATGCTATTTTGGAGGGA	360
QY	121	ValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysThrGlnLys	140
DB	361	GATGTTCAAAAGAACTTAAGATGATTCATCTTGAAGAAATTAAGTCTATCAAAAAG	420
QY	141	LysArgAspLysValAlaGluGluValLeuLeuAlaAlaLeuArgSerAspLeuAsnLys	160
DB	421	AAAGCTGATTAAGTTGGCGGAGAACTTCTTTTAAAGACATTCGTAAGCATCTTAAATGA	480
QY	161	SerValIleCysGluLysLysLysGluLysCysProValMetGlyArgGluSerAsp	180
DB	481	TCAGTCATATGTAAGAAAAAATTAAGATTTCCCTGTCATGGGAGGAGAAAGTAT	540
QY	181	GluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIleLeuLysLys	200
DB	541	GAGTTAAACAACCTGTGTCTGAACCAAGAAAGACATGTAAAGAAATTTTAAATGAAAAA	600
QY	201	AspLysLysCysGlyThrLeuLysThrAspValSerAlaAlaLeuGlySerPheLysLys	220
DB	601	GATAAGAAAGTGGGATCTTAAACAGATGTTTACAGACGCTAGGAAGTTTAAAAAA	660
QY	221	GluThrCysLeuGluLeuLeuGlnCysThrPheTyrlIleLysAsnCysGlyAspAsp	240
DB	661	GAAACATGCTTGAATTAAGTCAATGCTATTTTACATTTGAAATTTGGGAGACAC	720
QY	241	AspIleIleLysCysIleGluLeuGlnGlyLysCysGlnGluGlnAsnIleAlaIleMet	260

DB	721	GATTAATTAATGATTTGAATTTGGAGGAAAAAGCCCAAGAACAAATATGCTTAATG	780
QY	261	ProProGlyProAspPheAspProThrArgProGluAlaThrIleAlaGluAspIleGly	280
DB	781	CCACACGAGACCCGATTTTGAATCCATGACAGGCTACATATGACAGAGATATAGG	840
QY	281	LeuGluLysPheTyrlLysLysValGluGluAspGlyValPheIleGlyLysAsnIleLeu	300
DB	841	CTGAGAGGTTTAAAGAAAGTAGAGAGATGAGATTTTAAATTTGGAAGAAATCATCTA	900
QY	301	ArgAspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLysLysAsp	320
DB	901	AGAGATGCGACGCTTTGTGTGGCATTTGTTGATCCAGATTTCTAGCTTAAAAAAGAAC	960
QY	321	AspLysGluLysCysGluGluAlaLeuGlnLysSerCysLysAsnProHiseGluLys	340
DB	961	GACAAAGGAAATTCGAAAGAAAGCCCTTCAAAAGAGCTGCAAAAATCCCTCAATGACATAG	1020
QY	341	AlaLeuLysSerLeuCysLysLysAsnGlyLeuSerAsnAspGlyThrLysLysCysGlu	360
DB	1021	GCTTTAGAAAGTTTATGTATGAGAAAAATGCTTTAATGATGAGAACCAAAAAATGTGA	1080
QY	361	GluLeuGlnAsnAspIleAsnLysThrCysLysIlePheThrSerLysValThrAspAsn	380
DB	1081	GAAATTGCAAAATGATTTAAACAAACCTTGCAAAATTTCACTTCAAAAGTCACTAATAT	1140
QY	381	ArgLeuPheAspProThrLysGlyAsnAsnGluIleValGlyTrpGluGlyLeuProThr	400
DB	1141	CGTCTTTTGTATCCAAACAAAGGAAATATATGAAATTTGTGATGGGAAGGTTGGCAACA	1200
QY	401	PheLeuSerAsnGluAspCysAlaLysLeuGlnLysSerTyrlCysPheTyrlPheGluLys	420
DB	1201	TTTCTTTCGCAACGAAGATTTGTGCGAAATTTGAGTCCATATGTTTCTATTTTGAAGAAAA	1260
QY	421	CysProAspGlyGluAsnAlaCysLysAsnIleArgAlaThrCysTyrlLysArgGlyLeu	440
DB	1261	TGTCCAGATGAGAAATGCTGTAATTAATTAAGAGCAATGTTACAAAAGAGGACTT	1320
QY	441	AspAlaArgAlaAsnLysValLeuGlnLysAsnMetArgLysMetLysIleGlySerAsn	460
DB	1321	GATGACGGGCAATTAAGTCTCTCAAGAAATATGCGAGGAAATGTTATCAATGCTTCAAC	1380
QY	461	LysSerTrpLeuGlnLysPheGlnGlnGluLeuValCysGluLysLeuLysLys	480
DB	1381	AAAAAGCTGCTGTAAGTTTCAACGAATTAATAAGATATGAGAACTGAAAAA	1440
QY	481	GluAsnLysGlySerPheSerAsnAspGluLeuPheIleLeuCysValGlnProAlaLys	500
DB	1441	GAAACCAAGGAAATTTCTCAAGATGAATTAATTTCTGTGTGACAGCCAGCAAAA	1500
QY	501	AlaAlaArgLeuLeuThrHisAspLeuArgMetLysThrIlePheLeuArgGlnLysLeu	520
DB	1501	GCAAGCCGGTGTCTTACACATGATCTTCAATGAAATGAACTATCTTTTAAACAAACATG	1560
QY	521	AspGlnLysArgAspPheProThrAspLysAsnCysLysGluLeuGlyArgLysCysGln	540
DB	1561	GATCAAAAAGCGAAGTTTCCCAAGATTAATAAATGCAAGGAATTTGGGAGGAGAAAGTCCA	1620
QY	541	AspLeuGlyGlyAspSerLysGluIleThrTrpProCysHisSerTrpLeuGlnGlnLys	560
DB	1621	GATTTAGAGAGATTCTCAAAAGAAATTAATGATGATGATGATGATGATGATGATGATGATG	1680
QY	561	AsnArgLeuGlyThrThrGlnIleLeuLysGlnAlaIleLeuLeuAspGlnHisLysAspThr	580
DB	1681	AATGCTTTGGGAGATCAAGAAATTTTAAAGAGGTTTATTTGATGAAACAAAGATATCT	1740
QY	581	LeuLysAspGlnLysSerCysValLysTyrlLeuLysGluLysCysAsnLysTrpSerArg	600
DB	1741	TTGAAAGCCAAAGAAATTTGTGTAATAATCTTAATAAAGAAAGTGAATTAATGATGCTTGA	1800
QY	601	ArgGlyAspAspArgPheSerPheValCysValPheGlnAsnAlaThrCysGluLeuMet	620

[illegible]

Qy	981	lypbrophthlyscyethrtthnglyugluasapaspblaglyaspvallyvproseglu	1000
Db	2941	AAACCAACCAAGTGTATGACACGGAGAGAGAAAGATGATCAGAGACGTGAACCAAGTAC	3000
Qy	1001	Glyeuarfmetsergllytrpauvalmetcargglyvallylvalalmetvalileser	1020
Db	3001	GGCGTGAAGATGATGATGGCTGGAAATGTATGAGCGGGGCTGATAGTACGATGTTTCG	3060
Qy	1021	Phenectile 1023	
Db	3061	TTTCATGATT 3069	
RESULT 2			
US-10-654-416-11			
; Sequence 11, Application US/10654416			
; Publication No. US20040043409A1			
; GENERAL INFORMATION:			
; APPLICANT: Kovacs, et al.			
; TITLE OF INVENTION: Identification of a region of the major surface			
; TITLE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii			
; FILE REFERENCE: 4239-58054			
; CURRENT FILING DATE: 2003-09-02			
; PRIOR APPLICATION NUMBER: US/09/762,724			
; PRIOR FILING DATE: 2001-02-09			
; PRIOR APPLICATION NUMBER: PCT/US99/18750			
; PRIOR FILING DATE: 1999-08-17			
; PRIOR APPLICATION NUMBER: US 60/096,805			
; PRIOR FILING DATE: 1998-08-17			
; NUMBER OF SEQ ID NOS: 26			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 11			
; LENGTH: 3054			
; TYPE: DNA			
; ORGANISM: Pneumocystis carinii sp. f. hominis			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (1)..(3054)			
US-10-654-416-11			
Alignment Scores:			
Pred. No.:			
Score: 4398.50 Length: 3054			
Percent Similarity: 88.81% Matches: 838			
Best Local Similarity: 81.52% Conservative: 75			
Query Match: 81.29% Mismatches: 98			
DB: 19 Indels: 17			
Gaps: 9			
US-10-654-416-14 (1-1023) x US-10-654-416-11 (1-3054)			
Qy	2	AlaArgAlaVallysaRGlnAlaAlaGlyThrGlnmenSerIleAspGluGlu--His	20
Db	1	GGCGGGCGGTCACGCGACGAGTACACGAGACATCAGCGCCAAATGATGATGAAGTGCAT	60
Qy	21	ValLeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGlnGluCysLysLysLeu	40
Db	61	ATTTTGGCGTTGATTCTACCAAGAAATGCATATGTAAGATCAAAAATGCAAAAAAGTTTA	120
Qy	41	LysLysTrpCysGlnGluLeuThrGluAlaLysLeuSniIleGluGlnValHisArgLys	60
Db	121	GAATAAATACGCGAAGAGTTGAAAAAAGCATCATAGCATGGAATAAGTACATTAATAATG	180
Qy	61	LeuLysGlyPheCysGluAspGly-----LysAlaAspThrTrpLysCysLysGluLeu	77
Db	181	CTTAAGAATTTCTGTGGAAATGGGAAAGCAAGTAAAGCAATATCAAAATGTCAAGTCTTA	240
Qy	78	LysAlaAsnIleGluLysCysThrThrIleLys--GlyLysLeuLysGluAlaIle	96
Db	241	CAAGCAAAAGTACGGGGAAATGTACAAATTTTAAACACAAAAGCTAGGACACGCGTTA	300
Qy	97	LysLysLysIleGlnIleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeu	116

Dh 301 ACAAATCCA-----TCAGATGATAATTGCAAGAGAGTGAACGAACTGCCTA 348
Qy 117 PheLeuGluGlyValCysSerIysGluLeuLysAspAspCysAsnThrLeuAspLys 136
Dh 349 TTTTGGAGGGAGGATGC---CATTAATCTGTAGAAAGTTGTAACAACTAAGGATCTA 405
Qy 137 CysGlyGlnLysLysValAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSer 156
Dh 406 TGTATCCAGAAAAACGTGACGAGTGCAGAGAGTCTTTTGAGGGCCTTCGTAGT 465
Qy 157 AspLeuAsnGlySerValIleCysGluLysLysLeuLysLysIleCysProValMetGly 176
Dh 466 GATCTCAATAAACAAGAAACATGAAAAAACTGAAAGAGATTGGCCACTTGGAC 525
Qy 177 ArgGluSerAspGluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIle 196
Dh 526 AGGAAAGTATGATTAACGACCTTGTTGTAACCAAAAAAGACGTGGAGAAATTT 585
Qy 197 LeuIleGluLysAspLysLysCysGlyThrLeuLysThrAspValSerAlaLeuGly 216
Dh 586 ATAAAGAAAAAGATMAAAATGCACTACTTAAAGCAATGTTGCAACGACCTTGA 645
Qy 217 SerPheLysLysGluThrCysLeuGluLeuLeuGlnCysIlyrPheTyrlleGlyAsn 236
Dh 646 AGTTTAAAAAGAAATATGCTCTTGAAATTACTGAAACAATCTATTTTAACTGGAAAT 705
Qy 237 CysGlyAspAspAspIleIleLysCysIleGluLeuGlyLysCysGlnGluGlnAsn 256
Dh 706 TGGGAGACGACGATATATTAATGATTAATTGGAGGAAATGCAAGAACAAAC 765
Qy 257 ILeaIlyrMetProProGlyProAspPheAspProThrArgProGluAlaThrIleAla 276
Dh 766 ATTTTATATATACACACGACCCGATTTTGATCCAACTAGACAGAGCTACAC 825
Qy 277 GluAspIleGlyLeuGluLysPheTyrlsLysValGluGluAspGluValPheIleGly 296
Dh 826 GAGACATATGACCTGGATGACCTTTATAAAAAGCAGAGAGAGATGGCTTTTATTTGA 885
Qy 297 LysAsnHisLeuArgAspAlaThrAlaLeuLeuAlaLeuLeuIle---GlnAspSerSer 315
Dh 886 AAACATCATTTAAAGATGCACAGCTTTATTTGAGCTTTGTTAGAAAGATGATGATCA 945
Qy 316 LeuLysLysLysAspAspLysGluLysCysGluGluAlaLeuGlnLysSerCysLysAsn 335
Dh 946 GGAAGAAAAATATATATCGAGAAAAATGCAATAGATTCTGAAAGATTAATGCAAAAAAC 1005
Qy 336 ProHisGluHisGluAlaLeuGlnSerLeuCysLysLysAsnGlyLeuSerAsnAspGly 355
Dh 1006 TCTCAACGACATGAACTCTTAAAAATTTATGTAATTAATAGCTTAATGCATATGGA 1065
Qy 356 ThrLysLysCysGluGlnLeuGlnAsnAspIleAsnLysThrCysLysIlePheThrSer 375
Dh 1066 AAAGAAAAATGCAAGATTAAGAAAGATTTAAAAAACATGACAAACCTCMAACCA 1125
Qy 376 LysValThrAsnAsnArgLeuPheAspProThrLysGlyAsnAsnGluIleValGlyTrp 395
Dh 1126 ACGATTTCTTAAAAACCATCTTATGATCA-----AATGATAAATGTTGAGGCG 1176
Qy 396 GluGlyLeuProThrPheLeuSerAsnGluAspCysAlaLysLeuGluSerTyrlsCysPhe 415
Dh 1177 AGAAAACTGCACACTTTCTTAATATGAAAGCTGTGCAAGATTGGAAATCTTATTTGTTT 1236
Qy 416 TyrPheGluLysLysCysProAspGlyGluAsnAlaCysLysAsnIleArgAlaThrCys 435
Dh 1237 TACTACGAAAAAGCTGTGCAAAATGCCAAAGAGAGTGTATGATTAAGGCGACGCGT 1296
Qy 436 TyrlsValArgGlyLeuAspAlaArgAlaAsnLysValIleGlnGluAsnMetArgGlyMet 455
Dh 1297 TATATAGAGAGGCTTATGAGACGGGCAAAATAAAGCTGCAGAAAAATATCGTGGGTTA 1356
Qy 456 LeuHisGlySerAsnLysSerTrpLeuGlnLysPheGlnGlnGluLeuValLysValCys 475
Dh 1357 TTACGTGCTCAAAATCAAAAGTTGGCTTAAGGAGTTTCAACAAAGATTAAGTAAGTATG 1416

Qy 476 GluLysLeuLysLysGluAsnLysGlySerPheSerAsnAspGluLeuPheIleLeuCys 495
Dh 1417 AAGGAGCTA---AAAGAAAAATAAAGAAAGATTTCCCAAAACATGAAATATTTGTTCTGTGT 1473
Qy 496 ValGlnProAlaLysValAlaArgLeuLeuThrHisAspLeuArgMetLysThrIlePhe 515
Dh 1474 GTACAGCCACCAAAAGGTGACGATTACTTACACACATCATCAAAATGAGGGTTATCTTT 1533
Qy 516 LeuArgGlnGlnLeuAspGlnLysArgAspPheProThrAspLysAsnCysLysGluLeu 535
Dh 1534 TTACGACACAACTTGGATCAAAAGAGAGATTTTCCACACAGTATTAAGACTGCAAGAAATTA 1593
Qy 536 GlyArgLysCysGlnAspLeuGlyGluAspSerLysGluIleThrTrpProCysHisThr 555
Dh 1594 GGGAAAAAATGCCAAAGTTTAAAGAAAGATTCAAAAAATTAACATGGCCATGTCATACG 1653
Qy 556 LeuGluGlnGlnCysAsnArgLeuGlyThrThsGluIleLeuLysGlnValLeuLeuAsp 575
Dh 1654 CTGGAGCAGCAATGCATCCGCTGGGAGCTACAGAAATTTTAAAGCAGGTTTATGGAT 1713
Qy 576 GluHisLysAspThrLeuLysAspGlnLysSerCysValLysTyrlsLeuLysGluLysCys 595
Dh 1714 GAACACAAAGTACTTGTGAAGAGACCAAGAAAGTGTGTAATAATCCTTAAAGAAAGTGT 1773
Qy 596 AsnLysTrpSerArgArgGlyAspAspArgPheSerPheValCysValPheGlnAsnAla 615
Dh 1774 AATTAATGCTCTAGAAAGAGAGATGACCGTTTCTTTTGATGTGCTTCCAAACGCTT 1833
Qy 616 ThrCysGluLeuMetValLysAspValLysAspArgCysGluValPheLysLysAsnIle 635
Dh 1834 ACGTGTAGCTGATGATGTAAGAGCTGTAAGAGACGGGTGTGAAGTATTCAAAAAAATTA 1893
Qy 636 LysAlaSerTyrlleIleGluPheLeuGlnAsnAsnThrAsnLysIleThrThrLeuGln 655
Dh 1894 AAAGCTCATATATTTATGTAATTTCTTGAATAATTAACAATAATTAACAACCTGGA 1953
Qy 656 ArgAsnCysProSerTrpHisThrTyrlsCysAsnArgPheSerProAsnCysProGlyLeu 675
Dh 1954 AGAAATGCTCCCTCTTGGCATACCTATTTGCAATGATTTTCACTTAATTTGCCAGCTCTT 2013
Qy 676 ThrLysGluAsnSerCysThrLysIleLysLysHisArgGluProPheTyrlsArgLys 695
Dh 2014 ACGAAAGAGATGATTTACAAAATCAAGAGAGATGTGAGCCGTTCTATAAAAGAAAG 2073
Qy 696 AlaLeuGlnAspAlaLeuLysValGluLeuGlnGlyLysLeuThrAspLysSerLysCys 715
Dh 2074 GCCTTGAAGATGCTCCAAAGTAGACCTTCAAGAAATTTGACTGATTAATCTTAATGT 2133
Qy 716 GluProAlaLeuLysArgTyrlsCysThrValAlaGlyAsnValAsnAsnAlaSerIleSer 735
Dh 2134 GAACCTCATTTGAAAAAGATTTGTACAGTAGCGGAAACGTAAATTAATGCTCATCACT 2193
Qy 736 GlyLeuCysLysValAsnThrLysAspAsnSerGlyLysSerAspGluAspAlaArgLys 755
Dh 2194 GGGCTATGCAAGGTAAACCCAGAGATTACTTGAAAGAGTGTAGAGATGCTAGAAAG 2253
Qy 756 GluLeuCysGlnLysLeuValLysGluValGluGlnGlnCysLysAlaLeuProThrGln 775
Dh 2254 GAATCTGTGAGAAATTAAGTGAAGAAAGTGAAGAACAGTGCACAAAGCTTACCAACGAA 2313
Qy 776 LeuGlnGlnProAlaAlaAspLeuLysLysAspTyrlsThrTyrlsGluLysLysLys 795
Dh 2314 TTAGACAAACCGGACGCTGATTTAAAAAAGATTATGAACATATGAGAACTTAAGAAA 2373
Qy 796 ArgAlaGluGluAlaMetAsnLysSerSerLeuValLeuSerLeuIleLysLysAsnGlu 815
Dh 2374 CGTGCAAGAGAGCAATGAACCAAGTCCAGTCTGTGTTTGTCACTTAAACAAAAACAA 2433
Qy 816 SerAsnValSerLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGlyLeu 835
Dh 2434 AGTATGTATCAAAAAGTAAATAGCAAAAACAAGATTAAGAAATGCCGTTTCAACGGACTT 2493

QY 836 GlnaPThrThrLysHleValLysIleLeuArgGlyValLysAspValSerValThr 855
 DB 2494 CAAGATGACCAAAACATGCGAAATTAATCTACGAGAGAGATTAAGATGATCCGTACA 2553
 QY 856 GluLeuGluValAlaValAlaPheAspLeuAlaAlaGluValPheGlyArgTyrValAspLeu 875
 DB 2554 GAATTAGACATTAAGCATTTGATTTGGCAGACAGATATTGGAGATATGTGATTTG 2613
 QY 876 LysGluArgCysAsnLysLeuGluSerAspCysArgIleLysGluAspCysLysAspLeu 895
 DB 2614 AAGCAAGATGATTAATTAATGGAATCAGATTGCAAAATTAAGAGATGCGAAAGACTTA 2673
 QY 896 GluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluValLys 915
 DB 2674 GAAGAAATGACAAAAGATTAATTAAGGCTTGTGCAATCTGAAGCCCTGGAGCTGAG 2733
 QY 916 ProHleGluThrValIleArgLysSerThrThrThrThrThrThrThrThrValAla 935
 DB 2734 CCGCACCAAAACAGTACAGAAAGTACAAACGAACTACACACACAAACAAACCGTTGCC 2793
 QY 936 AspProLysAlaThrGluCysLysSerLeuGlnThrThrAspThrThrThrValIleThr 955
 DB 2794 GATCCGAAGGCAAGCAATGCAAAATCCCTACAGACAAACACACATGGGTACACAGACA 2853
 QY 956 SerThrIleThrSerThrSerThrIleThrSerThrIleThrSerLysIleThrLeuThr 975
 DB 2854 TCAGCACACACACACAGCTCTACATCACTACATCACTACATCACTACATCACTACATCA 2913
 QY 976 SerThrArgArgCysLysProThrLysCysThrThrThrGluGluAspAspAlaGlyAsp 995
 DB 2914 TCAACGAGGCGATCAAAACCAACCAAGTGTACGACAGG-----GATGATGACAGAAAG 2967
 QY 996 ValLysProSerGluGluLeuArgMetSerGlyThrAsnValMetArgGlyValIleVal 1015
 DB 2968 GTGAAGCCAAAGTGAAGGCTTGAAGGCTGCGAGTGAATGTGATGAGCGGCGTATGTA 3027
 QY 1016 AlaeMetValIleSerPheMetIle 1023
 DB 3028 GCATGCTTATTTCTGTTCAATGATT 3051

RESULT 3
 US-10-654-416-7
 ; Sequence 7, Application US/10654416
 ; Publication No. US20040043409A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kovacs, et al.
 ; TITLE OF INVENTION: Identification of a region of the major surface
 ; FILE REFERENCE: 4239-58054
 ; CURRENT APPLICATION NUMBER: US/10/654,416
 ; PRIOR APPLICATION NUMBER: US/09/762,724
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: PCT/US99/18750
 ; PRIOR FILING DATE: 1999-08-17
 ; PRIOR APPLICATION NUMBER: US 60/096,805
 ; PRIOR FILING DATE: 1998-08-17
 ; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 7
 ; LENGTH: 3084
 ; TYPE: DNA
 ; ORGANISM: *Pneumocystis carinii* sp. f. *hominis*
 ; FEATURES:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(3084)
 US-10-654-416-7

Alignment Scores:
 Score: 0
 Percent Similarity: 4352.00
 Best Local Similarity: 79.58%
 Length: 3084
 Matches: 826
 Conservative: 80
 Mismatches: 106

Query Match: 80.43% Indels: 26
 DB: 19 Gaps: 6
 US-10-654-416-14 (1-1023) x US-10-654-416-7 (1-3084)
 QY 1 MetAlaArgAlaValLysArgGlnAlaAlaGlyThrGlnAsnSerIleAspGluGlnHis 20
 DB 1 ATGCGCGCGCGGTCAACCGCGCAAAAGGTGCACAGAAATGCAATGATGTAGAGACAT 60
 QY 21 ValLeuAlaLeuLysLeuGluAspGlyLeuSerGluGlnGluCysLysLysLeu 40
 DB 61 GTTTAGCTTTGATTTTAAATAAATGATTTGAGATTCAGAAATGCAAAATGCAAACTAAGTTG 120
 QY 41 LysLysTyrCysGlnGlnLeuThrGluAlaLysLeuAsnIleGluGlnValHisArgLys 60
 DB 121 GAAGAAATTTGCAAAACATTAACAAATGCAAGATTAATCCAGAAAAAGTTTCAGAAAA 180
 QY 61 LeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLysAlaAsn 80
 DB 181 TTTAAAGATTTCTGTGATTAACGGGAAACGAAATGCAAAATGTCMAAGATCTTAAAAACAAA 240
 QY 81 IleGluLysLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLysIle 100
 DB 241 GTCATCAAAAATGCATTAATTTCAAGGAAACCTTCAACAGCTGCTAGAAAAAAATTT 300
 QY 101 GlnIleIleThrAspLysAspCysLysGluAsnGluGlnGlnCysLeuPheLeuGluGly 120
 DB 301 TCAGAAATTAACAGATGAGATTGCAAAAGATGAACAAATGCAATGCTTATTTTGGAGGA 360
 QY 121 ValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyrGlnLys 140
 DB 361 GCATGTCCAACAGAACTTAAGATGACTGCACATTAATTAAGAAATTAATCTTTTCAAAA 420
 QY 141 LysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeuAsnGly 160
 DB 421 GAACGAAACAATGTGAGAAAGAGTTCTTTGAGGCGCTTCTGTGATCTCAATGAA 480
 QY 161 SerValIleCysGluLysLysLeuLysGluIleCysProValMetGlyArgGluSerAsp 180
 DB 481 ACAAGACATGTGAAAAAACTGAAAGAGTTGGCCCGAAATTAAGAAAGAAAGCAT 540
 QY 181 GluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIleLeuIleGluLys 200
 DB 541 GAATTAACGAGGCTTGTCTTTATCAAAAAACAAACATGCGTAAGTCTTTTAACAAAA 600
 QY 201 AspLysLysCysGlyThrLeuLysThrAspValSerAlaAlaLeuGlySerPheLys--- 219
 DB 601 AAAAGTAAATGTGATCTCTTGAAAAAGAAAGTTGAAAGCACTTAAGAAATGAATTG 660
 QY 220 LysGluThrCysLeuGluGluLeuGluGlnCysTyrPheTyrIleGlyAsnCysGlyAsp 239
 DB 661 CGAAGAAAAATGCTACTATTACTTGAACATGTTACTTTTCAAGAGGAACTGTGAAGA 720
 QY 240 AspAsp-----IleIleLysCys 245
 DB 721 GACAAATCAAAAGTCATTAACCTAATATTAAGACTGCAAAAGAAATGTACAGAGTGT 780
 QY 246 IleGluLeuGlyLysCysGlnGlnAsnIleAlaIleTyrMetProProGlyProAsp 265
 DB 781 GATCAATTAACAGAAAGTGTGAAAAAAATTTGTTATTAAGCATCCAGATCCGAT 840
 QY 266 PheAspProThrArgProGluAlaThrIleAlaGluAspIleGlyLeuGluGluPheTyr 285
 DB 841 TTGCATCAACTAAGCCAGAGCTTACCTACAGAGACATAGGCTGGAAGACTTAT 900
 QY 286 LysLysValGluGluAspGlyValPheIleGlyLysAsnHisLeuArgAspAlaThrAla 305
 DB 901 AAGAGGAGAGAGATGATTTTGTGGAAGACAAACATGTAAAGATGCAACAGCT 960
 QY 306 LeuLeuAlaLeuLeuIleGlnAspSerSerLeuLysLysLysAspAspLysGluLysCys 325
 DB 961 TTGTTGACACTACTT-----CTTAAGAAACCCTTAAAAAAGAAAGATGT 1005

QY	326	GIUcLu1a1aLeuGluInLySeSeCybLybaenProhi8giu1a1aLeuGluSerLeu	345
Db	1006	ATAAAGCCCTTAATAAAAAAACCTGGCAAAACCTCATGAAACATGAGGCTTGAAAACTCA	1065
QY	346	CysLybLybaenGlyLeuSeSeRasnapsGlyThrybLybCy8giuGluLeuGlnAasnAsp	365
Db	1066	TGTAAAGAAAATTAAACCAAGTAGTGATGAACGAAAATATGATGAACCTGAAAAGAT	1125
QY	366	IleasnLybThrCybLybIlePheThrSerLybValThraenAasnArgLeuPheAasPro	385
Db	1126	GTTAAACAAACCTGTGAACAAGCTTACATCAACAATCTTAAAAACCGCTTACATTTC	1185
QY	386	ThrybGlybaenasnGluIleValcLyTrrpGluGlyLeuProthrPheLeuSeRasnGlu	405
Db	1186	CCTGATGGA-----ATTGCGGAATGGGGAAAATTACCGACATYCTTATGACATGAA	1236
QY	406	AAspCyba1aLybLeuGluSeSeThrCybPheThrybPheGlybLybCy8ProasnGlyGlu	425
Db	1237	GATTGTGCAAACTGAATCTTATGCTTATATTAATAAGAACTTGCCAGATGCAAA	1296
QY	426	Asn1a1aCybLybAsnIleArg1a1aThrCy8TrrLybAsArgGlyLeuAasPa1aArg1a1asn	445
Db	1297	GAAGCTGTATGAATGATGAGGGACGGCGCTTACAAGAGAAGGCTTGATGACGGGCAAC	1356
QY	446	LybValLeuGluInLybaenMetArgGlyMetLeuHi8giLySeRasnLySeTrrpLeuGlu	465
Db	1357	AGTGCTGTGCAAAAATAATATCGTGCGTTATTAACGCTGCTCAAAATCAAAATGGCTTAAG	1416
QY	466	LybPheGlnGluInLybValLybVala1aCy8GlybLybLeuLybLybGluAasnLybGly1Ser	485
Db	1417	GAGTTTCAACAAAGATTGAATGAATGAATGAAGAAGCTA---AAAGAAAATAAAGAAGT	1473
QY	486	PheSeRasnAasPgluLeuPheIleLeuCy8ValGlnProa1aLybVala1aArgLeuLeu	505
Db	1474	TTCCCAAAACGATGAATATATTGCTGTGTGTACACGACGAAAAGCTGCACGATTACTT	1533
QY	506	ThrHi8aPLeuArgMetLybThrIlePheLeuArgGlnInLybaenPgluLybArgAsp	525
Db	1534	ACACACGATCATCAATAGATGAGTTACTTTTACACAAACAATTGGATCAAAAGAGAGAT	1593
QY	526	PheProthrrAasPlybAasCybLybGluInLybArgLybCy8GlnAasPLeuGlyGluAasP	545
Db	1594	TTTCCGACAGATTAAGACTGCAGGACATGAGGAAAATATCCCAAGATTGAGAAAGAT	1653
QY	546	SerLybGluIleThrTrrProCy8Hi8ThrLeuGluGlnGlnInCy8AasArgLeuGlyThr	565
Db	1654	TCAAAAGAAATTACATGGCCATGTATACATCGAGACGACAAATGCAATCGCTGGGACCT	1713
QY	566	ThrgluIleLeuLybGlnValLeuLeuAasPgluHi8LybAasPThrLeuLybAasPgluInu	585
Db	1714	ACAGAAATTTTAACAGAGTTTATTTGATGAAACAAAGATTACTTGAAAGACCAAGAA	1773
QY	586	SerCy8ValLybTrrLybLeuLybGlybCy8AasnLybTrrpSeRArgArgGlyAasAasArg	605
Db	1774	AGTTGTGTAATATACATAAAGAAAAGGTATTAATATGCTTCAGAGAGGAGATGACCGT	1833
QY	606	PheSeRPhaValCy8ValPheGlnAasn1aThrCy8GlyLeuMetValLybAasPValLyb8	625
Db	1834	TTCTCTTTTGTATGTGTTTCCAAAACCGTACGTGTGACGTGATGTGTAAGAAGCTGAAA	1893
QY	626	AasPArgCy8GluValPheLybLybAasnIleLybA1aSerTrrIleGluPheLeuGlu	645
Db	1894	GACAGGTGTGAAGTATCAAAAATAATTAAGACCTCATATATTAATGAATTTCTTGAA	1953
QY	646	AasnAntThraenLybIleThrThrLeuGluAArgAasCy8ProSeTrrPheI8ThrTrrCy88	665
Db	1954	AATAATATCAAAATAAATAACAACACTGGAAAATAATTCCTCTTGCGCATACGATATGC	2013
QY	666	AasnArgPheSeProAasnCys8ProGlyLeuThrrLybGluAasN8Cy8ThrLybIleLyb8	685
Db	2014	AATAGATTTTCCCTTAATTTGCTCAGGCTTACGAAAGAGATATGTTGTACAAAATCCAG	2073
QY	686	LybHi8AArgGluProPheTrrLybAArgLybA1aLeuGluAasPa1aLeuLybVala1aGluLeu	705

Db	2074	AAGATTGTGAGCGCTTCTATAAAGAAAGGCGCTGGAGATCCTCTCAAAAGTAGACCTT	2133
QY	706	GIINGLYbysbuthrApblySerLysCySGluProAlaleuLybArgrLyrcySthrVal	725
Db	2134	CAAGGAAATTTGACTGTGATTAATCTTAATGTGAACCTGCAATGGAACAGATTTGTACAGTA	2193
QY	726	AlaGlybAnvalbAnbAnAlaSerLLeSerGlyLeuCybAlvalaAnThrLybApbAn	745
Db	2194	GGCGGAAACCTAAATATATATGCGTCAATCAGTGGCGCTTATGCAAACTAAACCAAGGATTAAC	2253
QY	746	SerGlyLysSerbPgluAspAlaArgLySGluLeuCybGluLysLeuValLysGluVal	765
Db	2254	TCGGAAGAGGTGATGAGATGCTGAGAAAGAACTCTGTAGAAATCAGTGAAGAAAGTG	2313
QY	766	GIINGluGInCysLybAlaleuProthGluLeuGlyGInProAlaAlaApbLeuLybLys	785
Db	2314	GAAGAACAGTGCAAAGACATTACCAACAGATTTAGACAAACCGGACAGCTGATCTTAAABAAA	2373
QY	786	AspLyLybThrTyrgGluLuleuLybLysArgAlaGluGluAlaMetbAnlybSerSer	805
Db	2374	GATTATTAAGCATATGTAGAGAACTTTAAGAAACGTGACAGAGAAACCAATGAACAAGTCCAGT	2433
QY	806	LeuValLeuSerLeuLileuLybLysbAnGluSerbAnValSerLybSerbAnSerLybAn	825
Db	2434	CTTGTTTTGTCACCTCATTAAGAAAAAGAAAGTAAGTATCAAAAAGTAATATACCAAAAC	2493
QY	826	LybAspLybAnAlaValSerbAnGlyLeuGInAspThrThrLybHisValLybLileu	845
Db	2494	AAGGATTAAGATATGCCGTTTCAAACGACCTTCMAAGTATCCACAAACATGTGAAAAATACTA	2553
QY	846	ArgArgGlyValLybAspValSerValThrgLuleuGluAlaLybAlaPheAspLeuAla	865
Db	2554	CGGAGAGAGAGTAAAGATGTATCCGTAAACAAATTAGAACCTTAACCATTTGATTGGCA	2613
QY	866	AlaGluValbPheGlyArgLyrrValAspLeuLybGluArgCybAnlybLysGluUserAsp	885
Db	2614	GCAGAAATATTGGAAATATGTAGATTGAAAGAAAGATGTATTAATTGGAATCAGAT	2673
QY	886	CysArgrLileuSGluAspCybLysAspLeuGluGluValCybLysLybLileuAnLybAla	905
Db	2674	TGCGAATTTAAGAGAGATTGCCAAAGACTTGAAGAAAGTATGCCAAAAAGATTAAAGAGCT	2733
QY	906	CysArgbAnleuLybProLeuGluValLybProHisGluThrValThrgLuserThrThr	925
Db	2734	TGTGCAATCTGAAGCCTCTGAGGTGAACCGGACGAACAGTACAGAAAGTACAACG	2793
QY	926	ThrThrThrThrThrThrThrValAlaAspProLybAlaThrGluCybLysSerLeu	945
Db	2794	ACAACTACAAACAAACAAACAAACCGTTGCCGATCCGAAAGCAACGCAATCAATCCTTA	2853
QY	946	GluThrThrAspThrTrpValThrgLInThSerThrHisThrSerThrSerThrLileThr	965
Db	2854	CAGCAACAGACACATGGGTTACACAGACATCCAGACACACACAGCAGCTTACTATACACA	2913
QY	966	SerThrLileThrSerLybLileThrLeuThrSerThrArgArgCybLysProThrLybCys	985
Db	2914	TCTACATCATCAATCAAAATTAACATTGACATCAACGAGCGAGTCAAAACCAACCAAGTGT	2973
QY	986	ThrThrGlyGluGluAspAspAlaGlyAspValLybProSerGluGlyLeuArgMetSer	1005
Db	2974	ACGACAGGCG-----GATGATGCGCAAGAGAGGTGAAGCAAGTGAAGGCTTGAGGGTGAAC	3027
QY	1006	GlyThrPbAnvalMetArgLybValLileValAlaMetValLileSerPheMetIle	1023
Db	3028	GGGGAATGTGATGAGAGGGGGGTGATGTAGCAATGGTATTTCGTTCAATGATT	3081
RESULT 4			
US-10-654-416-5			
; Sequence 5, Application US/10654416			
; Publication NO. US20040043409A1			
; GENERAL INFORMATION:			
; APPLICANT: Kovacs, et al.			


```
OY 779 Prolaalaapleuylvleasprtyrlsrthryrglugluuylvleasrglaaglu 798
DB 2350 GAGGAAAAAGACCTTAAAGATGATTTGAAACATTTGAAAACTTAAAAACAGGCGAAG 2409
OY 799 GluAlaMeCAnLysSerSerLeuValLeuSerLeuIleLysLysAsnGluSerAsnVal 818
DB 2410 AAAACAAATGAATTAATCCATCTTCTTTATCATCTGTTAAGAAAGATGAATAATACA 2469
OY 819 SerLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGlyLeuGlnAspThr 838
DB 2470 TCGAAAAAATAGACGAAAGACACAGATTAAGATACCGTTTCAACGCGCTTCAAGATACC 2529
OY 839 ThrLysHIsValLysIleLeuAsnArgGlyValLysAspValSerValThrGluLeuGlu 858
DB 2530 ACAGAACTATGAAATACTACGAGAGAGAGTTAAGATGATCCGTAAACGAATCTGAA 2589
OY 859 AlaLysAlaPheAspLeuAlaAlaGluValPheGlyArgTyrValAspLeuLysGluArg 878
DB 2590 GCTAAGGCAATTTGATTTGGTACAGATATTGGAAATATCTAGACTTGAAGAAAGA 2649
OY 879 CysAsnLysLeuGluSerAspCysArgIleLysGluAspCysLysAspLeuGluGluVal 898
DB 2650 TGTATTAATTCGATTCAGATTCAGAGATTAAAGAGATTGCAAGATTGTAAGAGAGTA 2709
OY 899 CysLysLysIleAsnLysAlaCysArgAsnLeuLysProLeuGluValLysProHIsGlu 918
DB 2710 TGTGAAAGATACAGAGATGTTTCGAAATTTAAACCTGAAGATGAAGCCGACGAA 2769
OY 919 ThrValThrGluSerThrThrThrThrThrThrThrThrThrThrValAlaAspProLys 938
DB 2770 ACAAGTGAAGAAAGACACAGACGACGACGACGACGACGACGACGACGACGACGACGACG 2829
OY 939 AlaThrGluCysLysSerLeuGlnThrThrAspThrTyrValThrGlnThrSerThrHis 958
DB 2830 GCAACAGATGCAAACTTTCAGACAAACAGATACATGATGATTAACACACTTCGACACAT 2889
OY 959 ThrSerThrSerThrIleThrSerThrIleThrSerLysIleThrLeuThrSerThrArg 978
DB 2890 ACCACACAGTCTACCATCATCTACATCATCATCAAAAATACACTCATCAACAAAGG 2949
OY 979 ArgCysLysProThrLysCysThrThrGlyGluGluAspAspAlaGlyAspValLysPro 998
DB 2950 CGTTCGCAACCAACCAAGTGTACGACAGG-----GATGATGACAGAGAGCTGAAGCG 3003
OY 999 SerGluGlyLeuAsnMetSerGlyTyrAsnValMetArgGlyValIleValAlaMetVal 1018
DB 3004 AGTAGGGAATGCAAGATGATGGTGAACCGTGAAGAGGGGGGTGATGATGATGATGAT 3063
OY 1019 IleSerPheMetIle 1023
DB 3064 ATTTGTTCAATGATT 3078

RESULT 6
US-10-654-416-1
; Sequence 1, Application US/10654416
; Publication No. US20040043409A1
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/10/654,416
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US/09/762,724
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3042
```

```
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3042)
US-10-654-416-1

Alignment Scores:
Pred. No.: 1,11e-240 Length: 3042
Score: 2762.50 Matches: 541
Percent Similarity: 69.10% Conservative: 177
Best Local Similarity: 52.07% Mismatches: 280
Query Match: 51.05% Indels: 41
DB: 19 Gaps: 18

US-10-654-416-14 (1-1023) x US-10-654-416-1 (1-3042)
OY 1 MetaLaArgAlaValLysArgGluAlaAlaGlyThrGlnAsnSerIleAspGluGluHis 20
DB 1 GTGGCGCGGCGGTTAAAGCGGACGTAACAGACGACGACGACGACGACGACGACGACGACG 57
OY 21 Val-----LeuAlaLeuIleLeuLysGluAspGlyLeuSerGluGluGluCysLysLys 38
DB 58 GTGGGCTTTTGGCTTTAATCTAATAAGATTTCTAAGAGATTAATAAAGATTAATAAAGGAA 117
OY 39 LysLeuLysLysTyrCysGluGluLeuThrGluAlaLysLeuAsnIleGluGluValHis 58
DB 118 AAATTTAGAAAACCTTCGAAAGATTAAGTGAACCAATCTAAGTCCAGAAACAAAGTACAT 177
OY 59 ArgLysLeuLysGlyPheCysGluAspGlyLysAlaAspThrLysCysLysGluLeuLys 78
DB 178 GAAAGTTTAAAGATTTCTGTGATACCAAAAACGCTGATTAATAAATGTAAGAACTAATAA 237
OY 79 AlaAsnIleGluLysCysThrThrIleLysGlyLysLeuLysGluAlaIleLysLys 98
DB 238 AAAATGTTGAAAAAATGCGGTGATTTTAAACAGATTAAGAAATGTTGTAAGAAAG 297
OY 99 LysIleGluIleIleThrAspLysAspCysLysGluAsnGluGluGluGluGluPheLeu 118
DB 298 GAACTTCAAAATTTGAAAAATGATGAGTGAACAAAATGAACACAGTCTGTTTAA 357
OY 119 GluGlyValCysSerLysGluLeuLysAspAspCysAsnThrLysAsnLysCysTyr 138
DB 358 GAAAGACATGCTCT---GATCTTACAAAGATTCGACAAAGATTTAAGAAACAAATGTTAT 414
OY 139 GlnLysLysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArg----- 155
DB 415 CAGATTAAGCGTGAATTAAGTGAAGAAAGTCTTTTAAGATAATTAATAAGAAAGAT 474
OY 156 ---SerAspLeuAsnGlySerValIleCysGluLysLysLeuLysGluIleCysProVal 174
DB 475 TTTAAAGATTAATAATTCGAC-----TGTGAAATAATTAACGTAAGTACTGTCAGAA 525
OY 175 MetGlyArgGluSerAspGluLeuThrAsnLeuCysLeuAsnGluLysGluThrCysLys 194
DB 526 TTAAGTCAAAATGATGACGATTAATGATGAATTAATGTTTGAATCAAAAAAATACGCTGAT 585
OY 195 AsnIleLeuIleGluLysAspLysCysGlyThr-----LeuLysThrAspVal 211
DB 586 AATCTTGAAGAAAGAAACGCAACAAAGTGAATCTTTCAAAAATCTTAAAAACGGAATTT 645
OY 212 SerAlaAlaLeuGlySerPheLys---LysGluThrCysLeuGluLeuLeuGluGluCys 230
DB 646 AAAACAAATTAAGAAAGATGAACCACTAAAAAATGCCCCATTAATTAATGAAGAAATGC 705
OY 231 TyrPheTyrIleGlyAsnCysGlyAspAspAspIleIleLysCysIleGluLeuGlyLys 250
DB 706 ATTTTTHATGATGAAGATGTGGAACGATTCGACGAGTGAAGATTTGGAAGAA 762
OY 251 LysCysGluGluGluAsnIleAlaTyrMetProGlyProAspPheAspProThrArg 270
DB 763 AAATGTCAAGAGAAATAATATTACTTACACATTATCATATCAAGGATTGATCTATACAA 822
```


Db 2926 TGTACGACGAGGAGGAAGATGACAGAGAGTGAACCGAGTGAAGGCTGAGGATG 29855
Qy 1005 SerGIYTPAaValMeArGIYValIleValAlaMeArValIleSerPheWetIle 1023
Db 2986 AGTGGGTGAGATGTATGATGAGAGGCTGTATTATGCAATGATGATTTTCATTCATGAT 3042

RESULT 7
US-10-654-416-3
; Sequence 3, Application US/10654416
; Publication No. US20040043409a1
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE REFERENCE: glycoprotein (MSG) gene of human Pneumocystis carinii
; CURRENT FILING DATE: 2003-09-02
; PRIOR FILING DATE: 2003-09-02
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3006
; TYPE: DNA
; ORGANISM: Pneumocystis carinii sp. f. hominis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3006)
US-10-654-416-3

Alignment Scores:
Pred. No.: 4,356-208 Length: 3006
Score: 2404.50 Matches: 479
Percent Similarity: 65.37% Conservative: 195
Best Local Similarity: 46.46% Mismatches: 320
Query Match: 44.44% Indels: 37
DB: 19 Gaps: 19

US-10-654-416-14 (1-1023) x US-10-654-416-3 (1-3006)

Qy 1 MetAlaArGAlaValIleArGIYValIleValAlaMeArValIleSerPheWetIle 20
Db 1 GTGGCGCGGCGGTCAAGCGCGGCTGCA--GCACAGAAATGATGTAAGAAAGAAATAT 57
Qy 21 ValIleuAlaLeuIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 40
Db 58 CTTTGGCTTGTGATTTTGAAGAAATGATGTAAGAAATGATGTAAGAAAGAAATAT 117
Qy 41 LysLysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 60
Db 118 AAAGAGTATTTGGAAGTTTAAAGAAATGTAAGAAAGAAAGAAAGAAAGAAAG 177
Qy 61 LeuLysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 79
Db 178 TTAGACGGAATCTGCAAGATGATGTAAGAAATGTAAGAAAGAAAGAAAGAAAG 237
Qy 80 AsnIleGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 99
Db 228 AAGGTTAAAGCAAAAGTATGTTTGAAGAAATGTAAGAAAGAAAGAAAGAAAG 297
Qy 100 IleGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 119
Db 228 GCTTCAACCTTGAAGATGATGTTTGAAGAAATGTAAGAAAGAAAGAAAGAAAG 357
Qy 120 GIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 139
Db 358 GGAGCATGTCAAGAACTTAAAGTAAATGTAAGAAAGAAAGAAAGAAAGAAAG 417
Qy 140 LysLysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 159

Db 418 AAAAAGAGAGCGATGACAGAAAGCTTTTAAAGATCTTAAAGGAACTTAAAG 477
Qy 160 GlySerValIleCysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 179
Db 478 GATAAAAACATGCAAAATTAAGTAAAGGCGGTTTGTCAAGAAATCAACAAAGAAAGT 537
Qy 180 AsnIleLeuLysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 199
Db 538 GATAGCTTAATTAAGTATGTTTGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 597
Qy 200 LysAspLysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 217
Db 598 AAAGATTAATCAATCAACCTTCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 657
Qy 218 PheLysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 237
Db 658 TTATTAAGAAAGAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 717
Qy 238 GIYAspAspIleIleLysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 257
Db 718 --GAAACAGATCGCAAGAGTGAAGAGTTTGAAGAAAGAAAGAAAGAAAGAAAGT 774
Qy 258 AlaTyrMetProPheGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 277
Db 775 GTTATGACAGACCAAGGTTTCACTTGTATCTTCAAGAAATTAAGAAAGAAAGAAAGT 834
Qy 278 AsnIleGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 297
Db 835 GAATAGACCTAAGAAAGTATGATGTAAGAAAGAAAGAAAGAAAGAAAGAAAGT 894
Qy 298 AsnIleLeuLysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 317
Db 895 CCATCAATTAATTAAGTGAAGTGTGTTTGAAGAAAGAAAGAAAGAAAGAAAGT 948
Qy 318 LysLysAspAspLysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 337
Db 949 CAATATCTTTAAAGATCAATGTAAGATGTTTAAAGAAAGAAAGAAAGAAAGT 1008
Qy 338 GIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 357
Db 1009 GAGCATATTAATTAAGAAAGTATTAAGTGAAGAAAGAAAGAAAGAAAGAAAGT 1068
Qy 358 LysCysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 377
Db 1069 AAAGCGAGAACTTAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1122
Qy 378 ThrAsnAspArgLeuPheAspProThrLysGIYValIleValAlaMeArValIleSerPheWetIle 397
Db 1123 AGGATTGAGAAATATTTGCTCCAGCT--AATGTAAGAAATTAATTTGCTGCAATATG 1179
Qy 398 LeuProThrPheLeuSerAsnGIYValIleValAlaMeArValIleSerPheWetIle 417
Db 1180 TTGCATACATTTCTTGGTGAAGAGAGTGAAGAAAGTGTGCGAGATGTTTAAATTTG 1239
Qy 418 GIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 437
Db 1240 AAAAGCCAAAGTCCA--CTTGAAGAAAGCCCTGCAATTAATTAAGAAAGAAAGAAAGT 1296
Qy 438 ArgGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 457
Db 1297 AAAGGCTTGAAGCAATGTAAGAAAGTGAAGAAAGTGAAGAAAGTGAAGAAAGTGAAG 1356
Qy 458 GlySerAsnLysSerTrpLeuGIYValIleValAlaMeArValIleSerPheWetIle 477
Db 1357 GCTTCAATTAAGATGCTTGAAGAAAGTGAAGAAAGTGAAGAAAGTGAAGAAAGTGAAG 1416
Qy 478 LeuLysGIYValIleLeuLysGIYValIleValAlaMeArValIleSerPheWetIle 497
Db 1417 ACAGAAAGAGAA--AGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1458
Qy 498 ProAlaLysAlaArgLeuLeuThrHisAspLeuArgMetLysThrIlePheLeuArg 517

Qy 373 PheThrSerIleValThrAsnAsnArgLeuPheAspProThrIleGlyAsnAsnGluIle 392
 Db 2038 -----CAAAGGATATT 2049
 Qy 393 ValGlyTrpGluGlyLeuProThrPheLeuSerAsnGluAspCys-----AlaLys 409
 Db 2050 CTAGCCCAAGAGAGCTGAAAAGCTTAT-----AAAGACTGCGTATCTCAAGCCAAA 2100
 Qy 410 LeuGluSerIleCysPheThrPheGluLysCysProAspGlyGluAsnAlaCysLys 429
 Db 2101 ACCGAACT-----GAGAAAAA-----GAATGCGAG 2127
 Qy 430 AsnIleArgAlaThrCysTrpLysArgGlyLeuAspAlaArgAlaAsnLysValLeuGln 449
 Db 2128 AATTTA-----CTACCCCTGAAGGAAAAAACTTTTGAA 2163
 Qy 450 GluAsnMetArgGlyMetLeuHisGlySerAsnLysSerTrpLeuGluLysPheGlnGln 469
 Db 2164 GAAGAAAGCCAAA-----GAGAGCGTTAAGCTTATTTGGATTGCGTATCTCAA 2211
 Qy 470 GluLeuVal-----LysValCysGluLysLeu-----LysLys 480
 Db 2212 GCCAAAACCGAAGCTGAGAAAAAAGAAATGCGAAATTTGCTCACCCCTGAAGCCAAAAA 2271
 Qy 481 GluAsnLysGlySerPheSerAsnAspGlyLeuPheIleLeuCysValGlnProAlaLys 500
 Db 2272 AAGTTAGAAGAAGCTAAAAAAGCTTAAAGCTTAACTTGATTCGTTCAAGAGCTAAG 2331
 Qy 501 -----AlaAlaArgLeuLeuThrHisAspLeuLysGlyMetLysThr 513
 Db 2332 AATGAAAAAGAAAAAAGAAATGCGAAGAAATTTGCTCACCCCTGAA-----GCGAAAAA 2385
 Qy 514 IlePheLeuArgGlnGlnIleuAspGlnLysArgAspPheProThrAsp-----Lys 530
 Db 2386 CTTTAAAGCAACAGCAAGCTAGATTGTTGAAAAACCTAAACCGATAAAGAAAGCAAAA 2445
 Qy 531 AsnCys-----LysGluLeuGlyArgLysCysGlnAspLeuGluLysAsp 545
 Db 2446 AAGCTTTGAAGAAGTCTCCCTAAAGACTTGCAGAAAAAGCTT-----TTAGCTAAAGAA 2499
 Qy 546 SerIleGluIleThrTrpProCys-----HisThrLeuGluGlnGln 559
 Db 2500 AGCGTTAAAGCTTAACTGATTTGCGTATCTCAAGCCAAAACCTGAAGCTGAAAAAAGAA 2559
 Qy 560 CysAsnArgLeuGlyThrThrGluIleLeuLysGlnValLeuLeuAspGlnHisLysAsp 579
 Db 2560 TGCAGAAATTAATCTCACCCCTGAAGCGAAGAA-----CTTTTGAAGAAAGCTTAAAAA 2613
 Qy 580 ThrLeuLysAspGlnGlnIleuSerCysValLysTrpLeu-----LysGluLys 594
 Db 2614 AGCGTTAAAGCTTAACTTGGATTGCGTATCTCAAGCCAAAACCTGAAGCTGAGAAAAAAGAA 2673
 Qy 595 CysAsnLysTrp-----SerArgArgGlyAsp 603
 Db 2674 TGGGAGAAATTAATCTCACCCCTGAAGCGAAGAACTTTTGAAGAAAGAAAGAGAGC 2733
 Qy 604 AspArgPheSerPheValCysValPheGln-----AsnAlaThrCys 617
 Db 2734 GTTAAACCTTAATCTGATTTGGTATCTCAAGCCAAAACGAAAGCTGAAAAAAGAAATGC 2793
 Qy 618 Glu-----LeuMetValLysAspValLysAspArgCysGluValPheLysLysAsnIleLys 636
 Db 2794 GAGAAATTTGCTCACCCCTGAATCGAAAAAAAGTTAAGAAAGAAAGAAAGCGTTAAG 2853
 Qy 637 AlaSerTrpIleIleGluPheLeuGluAsnAsnThrAsnLysIleThrThrLeuGluArg 656
 Db 2854 GCT-----TATTTGATTTGCGTATCTCAAGCCAAAACGAAAGCTGAGAA 2898
 Qy 657 AsnCysProSerTrpHisThrThrCysAsnArg-----PheSerProAsnCysProGlyLeu 675
 Db 2899 AAA-----GAATGCGAAAAAATTTGCTCAAGCGCTGAAGCGAAAAAACTT 2940
 Qy 676 ThrIleGluAsnSerCysThrLysIleLysLysHisArgGluProPheTrpLysArgLys 695

Db 2941 TTAGAGCAACAACGCGTAGATTGTTGAAAAAGCGTTAAACCGAAGCTGATTAAAAAAGC 3000
 Qy 696 AlaLeuGluAspAlaLeuLysValGluLeuGlnGlyLysLeuThrAspLysSerLysCys 715
 Db 3001 TGTGTCAAGAT---CTCCCTAAAGACTTGCAGAAAAAGGTTTAAAGCCAAAGAG----- 3051
 Qy 716 GluProAlaLeuLysArgTrpCysThrValAlaGlyAsnValAsnAsnAlaSerIleSer 735
 Db 3052 -----AGCTGAAGCTTAT----- 3066
 Qy 736 GlyLeuCysLysAlaAsnThrLysAspAsnSerGlyLysSerAspGluAspAlaArgLys 755
 Db 3067 -----AAAGACTGCGTATCAAAAGCTGAAATGAAAAAGAGAAA 3105
 Qy 756 GluLeuCysGlu-----LysLeuValLysGluValGluGlu 767
 Db 3106 AAGAAATGCGAAGAAATTAATCTCACCCCTGAAGCGAAGAAACCTTTAAGAAAGCTTAAAAA 3165
 Qy 768 GlnCysLysAlaLeuProThrGluLeuGlnProAlaAlaAspLeu-----LysLysAs 786
 Db 3166 AGCGTTAAGCTTAACTTGATTTGCGTATCTCAAGCCAAAACCTGAAGCTGAAAAAAGAA 3225
 Qy 786 PTrpLysThrTrpGluGluLeuLysLysArgAlaGluGluAlaMetAsnLysSerSerLe 806
 Db 3226 TGCAGAAATTAATCTCACCCCTGAAGCGA-----GAAAACCTTGAAGAAAGCTTAAAGAG 3279
 Qy 806 ValLeuSerLeuIleLys-----LysAsnGluSerAsnVal 818
 Db 3280 AGCGTTAACTTAAATATAAAGCTGCGTATCAAAAGCTGAAGATGAAAAAGAAAAAAGAA 3339
 Qy 818 LysLysSerAsnSerLysAsnLysAspLysAsnAlaValSerAsnGly-----LeuGln 836
 Db 3340 TGCAGAAATTAATCTCACCGCTGAAGCGAAGAAACCTTTAAGCAACAAGCTAGATTGT 3399
 Qy 836 AspThrThrLysHisValLysIleLeuAspArgGlyValLysAspValSerValThrGln 856
 Db 3400 TTGAAAAACGCTAAACCGAAGCTGATTAAGAAAAAGGTGTGCAAGATCTCCCTAAAG-GA 3458
 Qy 856 ValLeuGluAlaLysAlaPheAspLeuAlaAlaGluValPheGlyArgTrpValAspLeu-- 875
 Db 3459 CTTCAGAAAAAGGTT-----TTAGCTTAAAGAGAGCGTTAAGCTTATTTGACTGCGT 3512
 Qy 876 -----LysGluArgCysAsnLysLeu----- 882
 Db 3513 ATCAAGCTAGGAATGAAAAAGAAAGAAATGCGAAGAAATTTGCTCACCCCTGAAGC 3572
 Qy 883 -----GluSerAspCysArgIleLysGluAspCys----- 892
 Db 3573 GAAAAAACCTTTTGAAGAGAGCGAAGAGAGAGCTTAAAGCTTATTAAGACTGCGCTGTCA 3632
 Qy 893 -----LysAspLeuGluGluValCysLysLysIleAsnLysAlaCysArgAsnLeuLysPr 911
 Db 3633 AGCTAGAAATGAAGAGAA-----AGGAGAGCTTGCAGAAACTTACTCAC 3677
 Qy 911 GluGluValLysPro-----HisGluThrValThrGluSerThrThrThrThr 929
 Db 3678 GCGTGAAGCGAAGAACTTTAAGCAAGAAAGTTAAGAAAGCTTAAAGCTTATTTGGA 3737
 Qy 929 ThrThrThrThrValAlaAspProLys---AlaThrGluCysLysSerLeuGlnThr 948
 Db 3738 CTGCGTATCAAGAGCTAGAAATGAAAAAAGAAAGAAATGCGAAATTAATCTCACGCGC 3797
 Qy 948 rAsp 949
 Db 3798 TGA 3801

RESULT 10
 US-10-335-977-3219
 ; Sequence 3219, Application US/10335977
 ; Publication No. US20040052799A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGLAS SMITH et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 3219:
SEQUENCE CHARACTERISTICS:
LENGTH: 5460 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...5460
SEQUENCE DESCRIPTION: SEQ ID NO: 3219:
US-10-335-977-3219
Alignment Scores:
Pred. No.: 3 31e-17 Length: 5460
Score: 312.50 Matches: 248
Percent Similarity: 36.36% Conserved: 176
Best Local Similarity: 21.27% Mismatches: 413
Query Match: 5.78% Indels: 329
Gaps: 61
US-10-654-416-14 (1-1023) x US-10-335-977-3219 (1-5460)
QY 3 ArgAlaValLyAsrGlnAlaAlaGlyThrGlnAsnSerIleAspGluGlnHisValLeu 22
DB 1726 AAAGCTTAAAGACTGCGCATCTCAAGCCAAAGTGAAGCGTAAAGAAAGAAATGCCAG 1785
QY 23 AlaLeuIleLeuLyGlu--AspGlyLeuSerGluGlnGlyCysLyLeuLyLeuLyLeu 41
DB 1786 AAATTAACACAGCCGCGAAGCAAAACCTTTAGAGAGAAAGCCAAAGAGCGCTTAAG 1845
QY 42 LysTyrCysGlnGluLeuThrGlnAlaLysLeuAsnIleGluGlnValHisArgLyLeu 61
DB 1846 GCTTATTTGGATGCGTATCTCAGGCCAAAGTGAAGCTGAA----- 1887
QY 62 LysGlyPheCysGlu-----AspGlyLysAlaAspThrLysCysLyGluLeuLyLeu 78
DB 1888 AAAAAGATGCGAAGAAATGCTCACCTGAGCGAAGAAAGAAAGTTAGAGAGAGCCAA 1947
QY 79 AlaAsnIleGluLyLysCysThrThrIleLyGlyLyLeuLyGluAlaIleLyLeu 98

DB 1948 AAGAGCGTTAGAGCTTATTTGGATTGGCTATCAAAAGCCAAAGCGAAGCGTGAAGAA 2007
QY 99 LysIleGln---IleIleThrAspLysAspCysLyGluGlnGlnGlnCysLeu--- 116
DB 2008 GAATGCGAAGAAATTAATCTACGCGCTGAGCGAAGAAAGAAAGCTTTAGAGATCAAGCGCTAGAT 2067
QY 117 PheLeuGluGlyValCysSerLyGlyLeuLyAspAspCys-----Asn 131
DB 2068 TGTTGAAAAACGCTAAACCGATGAAGAACGAAAGAGTGTCTTAAGAAAGATCTCTTAA 2127
QY 132 ThrLeuArgAsnLys-----CysTyrGln 139
DB 2128 GACTTACAGAAAAAGTTTATGACCAAGAGCTTACGCTTATTTGGATTGGCTATCA 2187
QY 140 LysLyAsrGAsp-----LysValAlaGluGlnValLeuLeuArgAlaLeuArg 155
DB 2188 AAAGCCAAAAAGAGCTGAAGAAAGAAAGATGCGAAGAAATTCCTACCCCGAAGCGAGA 2247
QY 156 SerAspLeuAsnGlySerValIleCysGluLyLysLeuLyGluIleCysProValMet 175
DB 2248 AAGCTATTAGAGAGCTTAAGAGAGCGTTAAGCGCTTACAAAGAC---TGCCTTTTAA 2304
QY 176 GlyArgGluSerAspGluLeuThrAsnLeuCysLeuAsnGlnLyGluThrCysLyAsn 195
DB 2305 GCTAGGAAATGAAAAAGAG-----AAACAAGAAATGCGAGAAA 2340
QY 196 Ile-----LeuIleGluLyAspLyLysCysGlyThrLeuLyLeu 208
DB 2341 TTACTACGCGCTGAGAGGAGAAAGCTATTAGAGAAATCTAAGAAAGCGTTAAGCTTAT 2400
QY 209 ThrAsp--ValSerAlaAlaLeuGlySerPheLyLysGluThrCysLeuGluLeu 227
DB 2401 TTGGATTGCGATCAAAAGCCAAAGCAAGCTGAAGAAAGAAAGATGCGAAGAAATTCCTC 2460
QY 228 -----GluGln 229
DB 2461 ACCCTGAAGCGAAGAAAGCTATTAGAAAGAAAGCTTAAGAGCTTAAAGCTTACAAAGAC 2520
QY 230 CysTyrPheTyrIleGlyAsnCysGlyAspAspAspIleIleLysCysIleGluLeu 249
DB 2521 TGCGTATCAAGAGCTAGAGAT-----GAAAGAGAAACAGATGCGAAGAAATTA--- 2571
QY 250 GlyLyCysGlnGlnGlnAsnIleAlaTyrPheProGlyProAspPheAspProThr 269
DB 2572 -----CTC 2574
QY 270 ArgProGluAlaThrIleAlaGluAspIleGlyLeuGluGlnPheTyrLyLeuValGlu 289
DB 2575 ACGCTGAGAGGAGAAAGCTA-----TTAGAGAAATCTAAGAAAGCGTTAAG 2622
QY 290 Glu-----AspGlyValPheIleGlyLyAsnHisLeu-----ArgAspAlaThrAla 305
DB 2623 GCTTATTTGGATTCGATCAAAAGCCAAAGCAAGCTGAAGAAAGAAAGATGCGAGAAA 2682
QY 306 LeuLeuAla-----LeuLeuIleGlnAsp----- 313
DB 2683 TTGCTCACCCCTGAAGCGAGAAAGCTATTAGAAAGAAAGCTTAAGAGCTTAC 2742
QY 314 -----SerSerLeuLyLeuLyAspAspLyGlyLyCysGluGlnAlaLeu 329
DB 2743 AAAGACTGCGTATCAAGCTGAAGAAATGAAGAAAGAAAGAAAGATGCGAAGAAATTAATC 2802
QY 330 GlnLySerCysLyAsnProHisGluIleGluAlaLeuGlnGlnSerLeuCysLyLyAsn 349
DB 2803 ACGCTGAAGCGAAGAAAGCTATTAGAGAAATCAAGCGCTAGAT-----TGTGTAAGAAAC 2856
QY 350 GlyLeuSerAsnAspGlyThrLyLysCys---GluGluLeuGlnAsnAspIleAsnLys 368
DB 2857 GCTAAACCGAAGCTGAAGAAAGAAAGATGTCTCAAGATCTCCCTTAAGACTTACAGAAA 2916
QY 369 ThrCysAlaIlePheThrSerLyValThrAsnAsnArgLeuPheAspProThrLyGly 388

2917 -----AAGTTTTCGCAAGAGAGTGTAGGCTTATTGATTCGCTAACAAAGCC 2970
QY AsnAenGluLeuValGlyTTPGluGlyLeuProThrPheLeuSerAenGluAerCyAla 408
Db 2971 AAAAAGCAA-----GCTGAAGAAAAGAAATGCGAG 3000
QY LysLeuGluSerTyrCyPheTyrPheGluLysCyPProAerGly-----424
Db 3001 AAATTGCTCACC-----CCTGAAGGAGAAAGCTATT 3033
QY 425 GluAenAlaCyValysAenAlaLeuAla-----ThrCyTyrLysValGly-----439
Db 3034 GAAGAACTTAAGAGAGTGTAAAGCTTAACAAAGCTTCGCTAAGAGCTAGAAATGAA 3093
QY 440 -----LeuAerAlaArgAlaAenLysValLeuGlu 450
Db 3094 AAAGAGAAACAAGATCGAGAAATTACTCACCCTGAAGCTAGAAACTATTAGAGCA 3153
QY 451 AsnMetArgLysLeuLysGlySerAenLysSerTyrLeu-----464
Db 3154 GAAGTTAAA-----AAGAGCTTAAGCTTATTAGATCGCTTCAAGAGCT 3201
QY 465 -----GluLysPheGlnGlnGluValLysValCyValLys-----LeuLysGlu 481
Db 3202 AGGAATGAAAAGAGAAACAAGAA-----TCCGAGAAATTACTCACCCTGAA 3249
QY 482 AsnLysGlySerPheSerAenAerGluLeuPheLeuLysValGlnProAlaLys---500
Db 3250 GCGAGAAACTTTTAGAAGAAACAAGCTAGAT-----TGTTTGAAAACGCTAAAGCC 3303
QY 501 -----AlaAlaArgLeuLysThrLysAerLysAerLysThrLys 514
Db 3304 GAAGCTGAGAAAAGAGTGTGTCAAGAGATCTCCCTTAAGACTTCGAGAAAAGGTTT 3363
QY 515 -----PheLysArgLysGlnLysAer-----GlnLysArgAerPheProThr 528
Db 3364 GCTAAAGAGAGCTTAAGGCTTATTAGACTGCTATCAAGAGCTAGAAATGAAAAGAG 3423
QY 529 AsnLysAenCyValysGluLeu-----GlyTyrGlySerGlnAerLysGlu 544
Db 3424 AAAAAGAAATGCGAAGTGTGTCAAGCTGAGCGAGAAAACATTTAGAAAGATTTAA 3483
QY 545 AsnSerLysGluLysTyrProCy-----HisThrLeuGln 558
Db 3484 AAAAGCTTAAGCTTATTGATTCGCTATCAAAAGCCAAAAGAAAGCTGAAA 3543
QY 559 GlnCyAenArgLysGlyThrThrGlnLysLeuLysGlnValLeuLysAerGlnHisLys 578
Db 3544 GAATGCGAGAAATGCTCAGCTGAGCGAGAAAG-----CTATTAGAGAAAGCTTAA 3597
QY 579 AsnThrLeuLysAerGlnGluSerCyValLysTyrLeu-----LysGlu 593
Db 3598 GAGAGTGTAAAGCTTAACAAAGACTGCTATCAAGAGCTAGAAATGAAAAGAGAAACA 3657
QY 594 LysCyAenLysTyrPserAerArgLysAerAerPheSerPheValCyValPheGln 613
Db 3658 GAATGCGAGAA-----CTACTCACC 3678
QY 614 AsnAlaThrCyValLysLeuMetValLysAerValLysAerAerCyValLysPhe-----631
Db 3679 CCTGAAGCGAGAAACTATTAGACCAAGAACTTAAAGAGCTTAAGGCTTATTAGAC 3738
QY 632 -----LysLysAenLysLysAerLysLysGlu-----642
Db 3739 TGTGTATCAAGAGCTAGAAATGAAAAGAAACAAGATCGAGAAATTAAGTCAAGCTT 3798
QY 643 -----PheLysGluAenAenThrAenLysLysLysLysLysLysLysLysLys 658
Db 3799 GAAGCGAGAAATTTTAGAAGAAACAAGCGCAACAAGAAAGATTAAGCATTAAGATTC 3858
QY 659 ProSerTyrPheThrTyrCyAenAerPheSerProAen-----671
Db 3859 -----TTGAAAACGCGGATCTTAAGCAGACAGCGGCTATCATGAG 3900

672 ---CyAerGlyLeuThrLysGluAenSerCyThrLysLysLysLysLysLysLysLysLys 690
QY 3901 TGTTTGATGATGTTTGAAGCAATGAAGAGAAAGTCAAAATACCTGCAAGAAAGCGAGAA---3957
Db 691 PheTyrLysArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 710
QY 3958 -----AAGCTGTCTTGATTTGTTGAAAAGCGCTAGAACCGATGAA-----3999
Db 711 AsnLysSerLysCyValLysProAlaLeuLysAerGlyTyrSerThrValAlaGlyAenValAen 730
QY 4000 GAAAAGAGAAATGCCAA-----AACCTTATAGCGATTTGATTCAGAAATCCAA 4050
Db 731 AsnAlaSerLysSerGlyLeuCyValAenThrLysAerAenSer-----746
QY 4051 AATATA-----AAGCAGACAGAAACAATCAATTGAGTAAACA 4092
Db 747 -----GlyLysSerAerGluAerAlaArgLysGlu 756
QY 4093 GAAGGTTTCATCAAGCAAGCAAGCTGCTGATTAATTAATGATGACCTTACTGATCAAGAA 4152
Db 757 LeuCyValLysLysValLysGluValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 776
QY 4153 GCCATAGAGCAATGTTTAGAGGAGATTAAGCGATAGCGAAAGGCGCTTAATTTAGAAATC 4212
Db 777 GlyLysProAlaAlaAerLysLysAerLysLysAerLysLysLysLysLysLysLysLys 796
QY 4213 AAAGCAAGCTATGATGAAGTGAATGATCGATTTATGCGATCTAAGAAC---CGAAAGAC 4269
Db 797 AlaGluGluAlaAerLysLysSerLysLysLysLysLysLysLysLysLysLysLysLys 816
QY 4270 TTTGATTAACATGCGCGCTTAAGGTTATCCATTTGTTCAATGATTTTAAAAAGTGTGCG 4329
Db 817 AsnValSerLysSerAenSerLysAen-----LysAerLysAenAlaValSerAenGly 834
QY 4330 GATATTGCTACTATTAAAGCGCACTAAATGATGATGCGAGCAAAATAGCTAGCAATTCCT 4389
Db 835 Leu-----GlnAerThrThrLysHisValLysLysLysLysLysLysLysLysLys 849
QY 4390 ATTTATGCTTCATGAGAGCTGATTAATTAAGCAATTAAGCA---ACAGAGAAAGCAATT 4446
Db 850 LysAerValSerValThrGlnLysGlnLysLysAerAerAerAlaGluValPhe 869
QY 4447 AAGGATTAAGAT-----TTAGAGCTTAA-----TTAGCTTAAGGCTTTAGGT 4488
Db 870 GlyArgTyrValAer-----LeuLysGlu 877
QY 4489 GCGCATTAAGAAAGTGAAGTAAAGAAAGTAAAGAAAGTAAAGAAAGTAAAGCA 4548
Db 878 ArgCyAenLysLysGluSerAerCyAerGlyLysGluAerCyValysAerLysGlu 897
QY 4549 GAAGCAATTAAGATGACAAAGAT-----GTCCAGAAAGTCCAAATATATACAGCAA 4602
Db 898 ValCyValysLysLysLysLys-----AlaCyAerAenLysLysPro 911
QY 4603 ATGCTCTTAAGAAACAAGAAAGAAAGAGTGGGATTTTGAATGAATAATGTAATCC 4662
Db 912 LeuGluValLysProHisGluThrValThrGlnLysThrThrThrThrThrThrThr 931
QY 4663 ATTGATATTAAGAAAGAAAGAAAGAAAGAAAGATTAAGCAAGCTGCAACAGGCTTT 4722
Db 932 ThrThrValAlaAerProLysAlaThrGlnLysLysSerLysGlnThrThrAerThrTyr 951
QY 4723 ATAGCAAGATGATCC-----ACATTT 4746
Db 952 ValThrGlnThrSerThrHisThrSerThrSerThrThrThrThrThrThrThrThr 971
QY 4747 GTTTAGCAAAATGACCT-----CCCATTAATATCACTGATCTTAA 4791
Db 972 IleThrLysThrSerThr 977
QY 4792 GTAGATCCACTCTCAG 4809

```

RESULT 11
US-10-335-977-3220
/ Sequence 3220, Application US/10335977
/ Publication No. US20040052799A1
/ GENERAL INFORMATION:
/ APPLICANT: DOUGLAS SMITH et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
/ RELATING TO HELICOBACTER PYLORI FOR
/ DIAGNOSTICS AND THERAPEUTICS
/
/ NUMBER OF SEQUENCES: 10031
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: LAHIVE & COCKFIELD
/ STREET: 28 State Street
/ CITY: Boston
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02109-1875
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: CD-ROM ISO9660
/ COMPUTER: IBM PC Compatible
/ OPERATING SYSTEM: Windows NT 4.0
/ SOFTWARE: UNIX
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/335,977
/ FILING DATE: 30-Dec-2002
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/993,002
/ FILING DATE: 17-DEC-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Mandragouras, Amy E.
/ REGISTRATION NUMBER: 36,207
/ REFERENCE/DOCKET NUMBER: GTN-018
/
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)227-7400
/ TELEFAX: (617)742-4214
/
/ INFORMATION FOR SEQ ID NO: 3220:
/
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 5460 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: circular
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHEICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Helicobacter pylori
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (8) LOCATION 1...5460
/
/ SEQUENCE DESCRIPTION: SEQ ID NO: 3220:
US-10-335-977-3220

```

```

Alignment Scores:
Pred. No.: 4,08e-17 Length: 5460
Score: 311.50 Matches: 247
Percent Similarity: 36.54% Conservative: 179
Best Local Similarity: 21.18% Mismatches: 411
Query Match: 5.76% Indels: 329
DB: 19 Gaps: 61
US-10-654-416-14 (1-1023) x US-10-335-977-3220 (1-5460)
QY 3 ArgAlaValAlaLeuThrGlnAenSerIleAspGluGluHisValIleu 22
DB 1729 AAAGCTTAAAGAGCTGCGATCTCAAGCCAAAGCTGAAGTGAAGAAAGATGCGAG 1788
QY 23 AlaIleuIleuLeuGlu---AspGlyLeuSerGluGlnGluCyluLylyLeuLyly 41
DB 1789 AAATTACTCAGCGCTGAAGCCGAAAAAACTTTTGAAGAAAGAAAGCCAAAGAGCGCTTAAG 1848
QY 42 LyuTyGyGlnGluLeuThrGlnAlaLylyLeuAenIleGluGlnValHisArgLylyLeu 61
DB 1849 GCTTATTGGAATGCGTATCTAGGCGCAAAAGCTGAAGCTGAA-----1890

```

```

QY 62 LyuGlyPheCyGlu-----AspGlyLysAlaAspThrLyCyLyGluLeuLy 78
DB 1891 AAAAAAGATGCGAAGATTCCTACACCTGACGAGAAAAAGTTAGAGAAAGCCAAA 1950
QY 79 AlaAenIleGluLylyCyThrThrIleGlyLylyLeuLyGluAlaIleLyly 98
DB 1951 AAGAGCGTTAGAGCTTATTGGATTCGATCAAAAGCCAAAGAAAGCTGAAGAGAAA 2010
QY 99 LysIleGln---IleIleThrAspLysAspCyLyGluAenGluGlnCysLeu--- 116
DB 2011 GAATGCGAAGAAATTACTCAGCGCTGAAGCCGAAAAAACTATTAGAAATCAAGCGCTAGAT 2070
QY 117 PheLeuGluGlyValCySerLyGluLeuLysAspCyS-----Asn 131
DB 2071 TGTTTGAAAAACGCTAAACCGATGAAGAAAGAGTGCTTGAAAGATCTCCCTAAA 2130
QY 132 ThrLeuArgAenLyS-----CysTyGln 139
DB 2131 GACTTACAGAAAAAGTTTATGCCAAAGAGCTTAGCGTTTATTGGATTCGATCA 2190
QY 140 LysLyAspArgAsp-----LysValAlaGluGluValLeuLeuArgAlaLeuArg 155
DB 2191 AAAGCCAAAGAAAGAGCTGAAGAAAGAAAGATGCGAAGATTCACCCCTGAAGCGAGA 2250
QY 156 SerAspLeuAenGlySerValIleCyGluLylyLeuLyGluIleCySProValMet 175
DB 2251 AAGCTATTAGAAAGAGCT-----AAGAAAGCGCTTAAAGCTTACAAAGACTCGCTT 2301
QY 176 GlyArgGluSerAspGluLeuThrAsnLeuCySLeuAenGlnLyGluThrCyLySAsn 195
DB 2302 TCAGAGCTAGAGATGAA-----AAAGAAACAAAGATGCGAGAAA 2343
QY 196 Ile-----LeuIleGluLySAspLySlyCyGlyThrLeuLyS 208
DB 2344 TTACTCAGCGCTGAAGCGAAGAACTATTAGAGAAATCTAAAGAAAGCTTAAAGCTTAT 2403
QY 209 ThrAsp---ValSerAlaAlaLeuGlySerPheLySlyGluThrCySLeuGluLeu 227
DB 2404 TTGATTCGCTATCAAAAGCCAAAGCAAGACGTGAAGAAAGATGCGAAGATTCGTC 2463
QY 228 -----GluGln 229
DB 2464 ACCCTGAAGCGAAGAAAGCTTATGAAGAGCTTAAAGAGTGTAAAGCTTACAAAGAC 2523
QY 230 CysTyGlyPheTyIleGlyAenCyGlyAspAspAspIleIleLyCySleGluLeuGly 249
DB 2524 TGCGATCAAGAGCTAGAGAT-----GAAAGAGAAACAAGATGCGAAGAAATTA--- 2574
QY 250 GlyLyCySglnGluGlnAenIleAlaTyMetProGlyProAspPheAspProThr 269
DB 2575 -----CTC 2577
QY 270 ArgProGluAlaThrIleAlaGluAspIleGlyLeuGluGluPheTyLyLyValGlu 289
DB 2578 ACGCTGAGGAGGAGAACTA-----TTAGAAAGATCTAAAGAAAGCTTAAAG 2625
QY 290 Glu-----AspGlyValPheIleGlyLysAenHisLeu-----ArgAspAlaThrAla 305
DB 2626 GCTTATTGATTCGATCAAAAGCCAAAGCAAGCTGAAGAAAGAAATGCGAGAAA 2685
QY 306 LeuLeuAla-----LeuLeuIleGlnAsp-----313
DB 2686 TTGCTCACCCCTGAAGCGAGAAAGCTATTAGAAAGAGCTTAAAGAGAGTGAAGCTTAC 2745
QY 314 -----SerSerLeuLylyLySAspAspLySlyGluLyCyGluGluValLeu 329
DB 2746 AAAGATCGCGTACAGAGCTAGAAATGAAAGAAACAAAGATGCGAAGAAATTTACTTC 2805
QY 330 GlnLySerCyLySAsnProIleGluIleGluAlaLeuGluSerLeuCyLyLySAsn 349
DB 2806 ACGCTGAAGCGAAGAAAGCTATTAGAGATCAAGCGCTAGAT-----TGTGTAAGAAAC 2859

```

Qy 350 GlyLeuSerAenbpglyThrLyLeuCyS---GluGluLeuGlnAenApIleAenLyS 368
Db 2860 GCTAAACCGAAGTGAAGAAAAAGGTGTCTCAAGATCTCCCTAAAGCTTACAGAAA 2919
Qy 369 ThrCyblysllePheThrSerLyValThrAenbAenrgLeuPheAenProThrLyGly 388
Db 2920 -----AAGCTTTAGCCAAAGAGAGTGTAGGCTTTATTGTTGCTGATCAAAAGGCC 2973
Qy 389 AenbAenGluLeuValGlyTPGluGlyLeuProThrPheLeuSerAenGluAenCybala 408
Db 2974 AAAACGAA-----GCTGAAGAAAAGAAATGGGAG 3003
Qy 409 LyLeuGluSerTyCybPheTyPheGluLybCySProAenGly-----424
Db 3004 AAATTGCTCAC-----CCTGAAGGAGAAAGCTATT 3036
Qy 425 GluAenAlaCybAenIleAenGlyA-----ThrCybTyLybAenGly-----439
Db 3037 GAAGAACTAAAGAGCTTAAAGCTTAAAGAGCTGCTATCAAGAGCTAGAGATGAA 3096
Qy 440 -----LeuAenAlaAenrgAlaAenLybValLeuGlnGlu 450
Db 3097 AAAGAGAAACAAGATCGAGAAATTAAGCTACGCTGAAAGCTAGAGAACTATTAGAGCA 3156
Qy 451 AenMeAenrgGlyMeLeuHleGlySerAenLybSerTyPheLeu-----464
Db 3157 GAAGTTAAA-----AAGAGCTTAAAGCTTATTAGACTGCTGCTTCAAGAGCT 3204
Qy 465 -----GluLybPheGlnGlnGluLeuValLybValCybGluLyS---LeuLybLyGlu 481
Db 3205 AGCAATGAAAAAGAGAAACAAGAA-----TGCAGAAATTAAGCTACAGCCCTGAA 3252
Qy 482 AenLybGlySerPheSerAenAenpGluLeuPheIleLeuCybValGlnProAlaLyS---500
Db 3253 GCGAGAAACTTTTAGAAGAACTCAAGCCCTAGAT-----TGTGGAAGAAACGCTTAAACC 3306
Qy 501 -----AlaAlaAenrgLeuThrHisAenrgLeuAenrgMeLybThrIle 514
Db 3307 GAAGCTGAGAAAAAGAGTGTCTCAAGATCTCCCTAAAGCTTGCAGAAAAAGCTTTTA 3366
Qy 515 -----PheLeuAenrgGlnGlnLeuAenP-----GlnLybAenrgAenPheProThr 528
Db 3367 GCTAAAGAGAGCTTAAAGCTTATTAGACTGCTATCAAGAGCTAGAAATGAAAAAGAG 3426
Qy 529 AenLybAenCybLybGluLeu-----GlyAenrgLybCybGlnAenrgGluGlu 544
Db 3427 AAAAAGAAATGCGAAGATGTCTCAAGCTGAAAGCAAGAACTATTAGAAAGATCTAAA 3486
Qy 545 AenPheTybGluIleThrTyPheProCyS-----HisThrLeuGluGln 558
Db 3487 AAAAGCTTAAAGCTTATTGATTTGCTATCAAAAGCGCAAAAGCAAGCTGAAAAAAA 3546
Qy 559 GlnCybAenAenrgLeuGlyThrThrGluIleLeuLybGlnValLeuLeuAenrgIleHisLyS 578
Db 3547 GAATGCGGAAATTTGCTCACAGCTGAAAGCAAGAAAG-----CTATTAGAAAGAGCTTAA 3600
Qy 579 AenPheThrLeuLybAenrgGlnGluSerCybValLybTyLeu-----LybGlu 593
Db 3601 GAGAGCTTAAAGCTTAAAGAGCTGCTATCAAGAGCTAGAAATGAAAAAGAAACAA 3660
Qy 594 LybCybAenLybTyPheSerAenrgGlyAenAenrgPheSerPheValCybValPheGln 613
Db 3661 GAATGCGGAAATTTGCTCACAGCTGAAAGCAAGAAAG-----CTACTCAC 3681
Qy 614 AenAlaAenrgGlyLeuMeValLybAenrgValLybAenAenrgCybGluValPhe-----631
Db 3682 CCTGAAGCGAAGAACTATTAGAGCAAGAGTAAAGAGCTTAAAGAGCTTATTAGAC 3741
Qy 632 -----LybLybAenIleLybAenAenrgTyIleIleGlu-----642
Db 3742 TGTGTATCAAGAGCTTAAAGAAATGAAAGAAACAAGATCGGAAATTAAGTCAAGCCCT 3801
Qy 643 -----PheLeuGluAenbAenThrAenLybIleThrThrLeuGluAenrgAenCyS 658

Db 3802 GAAGCGAGAAATTTTAGAGAAACAGCGCAACAAAGATAAAGCATAAAGATATTC 3861
Qy 659 ProSerTybHisThrTyCybAenAenrgPheSerProAen-----671
Db 3862 -----TTGAAAAACCGGATCTTAAACAGACAGCGGCTATCATAGAG 3903
Qy 672 -----CybProGlyLeuThrLybGluAenSerCybThrLybIleLybHisAenrgPro 690
Db 3904 TGTGTGATGTGTTGAGCGATGAAGAGAAAGTCAAAATCACTGCAAGAAAGCAGAGAA---3960
Qy 691 PheTyLybAenrgLybAlaLeuGluAenrgAlaLeuLybValGluLeuGlnGlyLybLeuThr 710
Db 3961 -----AAGGCTGCTTGATGATTTGAAAAACGCTAGAGACCGATGAA-----4002
Qy 711 AenLybSerTyCybGluProAlaLeuLybAenrgTybTybThrValAlaGlyAenValAen 730
Db 4003 GAAAAAGAAATGCCAA-----AACCTTAAAGCATTTGATCAAGAAATCCAA 4053
Qy 731 AenAlaSerIleSerGlyLeuCybLybAlaAenThrLybAenAenSer-----746
Db 4054 AATTA-----AAGCAACAGAACAAATCAATTAAGTAAACA 4095
Qy 747 -----GlyLybSerAenrgGluAenrgAlaAenrgLybGlu 756
Db 4096 GAAGGTTGCATCAAGCAAGAGTCTGATTAATTAAGATGACCTTACTGATCAAGAA 4155
Qy 757 LeuCybGluLybLeuValLybGluValGluGlnGlnCybLybAlaLeuProThrGluLeu 776
Db 4156 GCCATAGAGCAATGTTTAGAGGAGATTGAGCATGACGAAGGCGCTTAAATTTAGAGATC 4215
Qy 777 GlyGlnProAlaAlaAenrgLybLeuAenrgTybTybThrTybGluGluLeuLybAenrg 796
Db 4216 AAAGCAAGCTATGATGAAGTGCATGCTGATTTATAGCATCTAGAAC-----CGAAAAAC 4272
Qy 797 AlaGluGluAlaAenAenLybSerSerLeuValLeuSerLeuIleLybLybAenrgLyb 816
Db 4273 TTTGATTAACATGGCGGCTTAAAGTTATCAATGTTTCAATGATGATTTAAAAAGTGTGC 4332
Qy 817 AenValSerLybSerAenSerLybAen-----LybAenrgLybAenAlaAenAenrgLyb 834
Db 4333 GATATTTGCTACTTAAAGCGCACTAATGTGATGCGGACAAATAGCTAGCATTAAGCTT 4392
Qy 835 Leu-----GlnAenPheThrLybHisValLybIleAenrgAenrgGlyVal 849
Db 4393 ATTATATCTTCATAGAGCTGATATTACTTAAGCAATACAA-----ACAGAGAAACCATT 4449
Qy 850 LybAenrgValSerValThrGluLeuGluAlaLybAlaPheAenrgLeuAlaGluValPhe 869
Db 4450 AAGCATTAAGAT-----TTAGAGCTTAA-----TTAGCTAAGGCTTTAGGT 4491
Qy 870 GlyAenrgTybValAenP-----LeuLybGlu 877
Db 4492 GCGCATTAAGAAAGTGAAGTAAAGAAAGTAAAGAAAGTAAAGCAAGAAATCAAGCA 4551
Qy 878 ArgCybAenLybLeuGlnGluSerAenrgCybAenrgIleLybGlnAenrgCybLybAenrgGluGlu 897
Db 4552 GAAAGCAATTAAGATGACAAAGAT-----GTCCAGAAAGTGCACAAATATATAGCGAA 4605
Qy 898 ValCybLybLybIleAenLyS-----AlaCybAenrgAenrgLeuPro 911
Db 4606 ATGCTCTTAAAGAAACAAAAAGAAAGAGTGGGATTTTGAAGTGAAGAAATGGAATCCC 4665
Qy 912 LeuGluValLybProHisGluThrValThrGluSerThrThrThrThrThrThr 931
Db 4666 ATTGATATTAAGAAAGAAAGAAAGAAACAAGATTAACAAGCCCTGTCAACAGAGCTTT 4725
Qy 932 ThrThrValAlaAenrgProLybAlaThrGluCybLybSerLeuGlnThrThrAenrgThrTyP 951
Db 4726 ATAGGCAAGATGATCCC-----ACATTT 4749
Qy 952 ValThrGlnThrSerThrHisThrSerThrSerThrIleThrSerThrLyS 971


```

Db      2722 AGAAACTTTAGAGAGAGCTTAAAGAGCGTTAAAGCTTAAAGC----- 2769
Qy      493 ILeuCyValGlnProAlaLys-----AlaAlaArgLeuLeu 505
Db      2770 -----TGCCTATCAAAAGCTAAGATGAAAGAAAAAGAAATGCGAGAAATTACTC 2823
Qy      506 ThrHisPheLeuArgMetLysThrIlePheLeuArgGlnGlnLeu----- 520
Db      2824 ACCGCTGAAGCGAAAAA-----CTTTAGAGCAACAAGCTGATGTTGTTGAA 2874
Qy      521 -----AapGlnLysArg-----AapPheProThrAapLysAsn 531
Db      2875 AACGCTAAACCGAAGCTGATTAAGAAAAAGGTGTCTCAAAAGATCTCCCTAAAGACTTCAG 2934
Qy      532 CybLysGlnLeuGlyArgLys-----CybGlnAapLeuGlnGlu 544
Db      2935 AAAAAGGTTTATGCTTAAAGAGAGCGTTAAGGCTTATTTGACTGCTGATCAAGAGCTAAG 2994
Qy      545 AapSerLysGlnIleThrTrpProCysHisThrLeuGlnGlnGlnCybAapArgLeuGly 564
Db      2995 AATTAAAAAAGG-----AAAAAAGATGCGAAGAAATTGCTC 3030
Qy      565 ThrThrGlnIleLeuLysGlnValLeuLeuAapGlnHisLysAapThrLeuLysAapGln 584
Db      3031 ACCCTGAAGCGAAAAA-----CTTTAGAGAAAGCGAAAGAGAGCTTAAAGCTTAT 3084
Qy      585 GluSerCybValLysTrpLeuLysGlnLysCybAapLysTrpSerArgArgGlyAapAsp 604
Db      3085 AAAAGACTGC-----CTCTCTCAAGCTAAGAAATGAAAGAAAGAGAG----- 3126
Qy      605 ArgPheSerPheValCybValPheGlnAapIleThrCysGln-----LeuMetValLysAap 623
Db      3127 -----GCTTCCGAAACTTACTACTACCCCTGAA 3153
Qy      624 ValLysAapArgCybGlu-----ValPheLysLysAapHisLysAalaserTrpIle----- 640
Db      3154 GCGAGAAAGCTTATGAGACAGAAAGTTAAGAAAGACTTAAAGCTC-----TATTTGAGACTGC 3210
Qy      641 IleGlnPheLeuGlnAapHisThrAapLysIleThrThrLeuGlnLysAapCysProSer 660
Db      3211 GTATCAAGAGCTAAGAAATGAAAGAAAGAAAGAAATGCGAAGAAATTACTACCCCTGAA 3270
Qy      661 TrpHisThrTrpCysAapArgPheSerProAsnCys----- 672
Db      3271 GCGAGAAATTTTATGAGAGCAAGCTGCTAATTTGTTGAAAAAGCTCGAAGATGAAAGAA 3330
Qy      673 -----ProGlyLeuThrLysGlnAapSerCysThrLys 683
Db      3331 GAAAGAAAGCATGCTTAAATATCTCCCTAAAGACTTACAGAAATATTTTATGCTTAA 3390
Qy      684 -----IleLysLysHisArgGlnProPheTrpLys----- 693
Db      3391 GAGAGCTTAAAGCTTATTAAGAGCTGCTCTCTCAAGCTAAGAAATGAAAGAAAGAGAA 3450
Qy      694 -----ArgLysAlaLeuGlnAapAlaLeuVal 703
Db      3451 GCTTGCAGAAACTTACTACGCTCGAAGCGAAGAACTTAAAGCAAGAAAGTTAAGAA 3510
Qy      704 GlnLeuGlnGlnLysLys-----ThrAapLysSerLys 714
Db      3511 AGCGTTAAGGCTTATTTGAGACTGCGTATCAAGAGCTAAGAAATGAAAGAAAGAAAGAA 3570
Qy      715 CybGln-----ProAlaLeuLysArgTrpCysThrValAlaGlyAapValAsn 730
Db      3571 TGCAGAAATTTACTACGCGCTGAAGCGAAGAAATTTTAAAGCAAGAACTTCAACAAAA 3630
Qy      731 AapAlaSerLysSerGlyLysCybLysAlaAapThrLysAapAsn----- 745
Db      3631 GATTAAGCGATCAAAAGAT-----TGCCTGAAAAAGCCGATCTTAAAGCAGAGCGGCTATC 3687
Qy      746 -----SerGlyLysSerAapGlnAapAlaArgLysGlnLeu----- 757

```

```

Db      3688 ATGAAGTTTGGATGCTTGAAGCATGAAGAGAGCTCAAAATCTTCAGAAAGCTAGA 3747
Qy      758 -----CybGlnLysLeuValLysGlnValGlnGlu-----GlnCys 769
Db      3748 GAAAGCGCTGTGGGATTTGTTGGCTATGCTTAAACCGATGAAGAAAAAGAAATGC 3807
Qy      770 LysAlaLeuProThrGlnLeuGlnGlnProAla----- 780
Db      3808 CAAGAACTTATAGCGATTGATTCAGAAATCCAAATTAAGAAAGACAAACAAACAA 3867
Qy      781 -----AlaAsp 782
Db      3868 AATCAATTGATTAACAGAAAGGTTGCATCAAGCAAGAGCTGTTGATTAAGCTTAT 3927
Qy      783 LeuLysLysAapTrpLysThrTrpGlnGlu----- 792
Db      3928 GACCTTACTGATCAAGAGCGCTTGAAGCAATGTTTAAAGGCTTGAAGCATGTAAGG 3987
Qy      793 -----LeuLysLysArgAlaGlnGlnAla----- 800
Db      3988 GCGCTAATTCTAGGAATTAACAGCAAGAGCTGATGAAGTGTGATTTATAGCATCTA 4047
Qy      801 MetAsnLysSerSerLeu-----ValLeuSerLeuIle 811
Db      4048 AGAAACCGTAAACCTTTGATTAACATGCGCGCTAAGAGTTATCATTCATTTCAACAAATGAT 4107
Qy      812 LysLysAapGlnLysAapValSerLysSerAsnSerLysAsn-----LysAapLysAsn 829
Db      4108 TTCAAAATAGCGCGGATTTGCCATATTAAAGCCCTAATGTTGATGCGGAGAAATA 4167
Qy      830 AlaValSerAsnGlnLys-----GlnAapThrTrpLysHisValLysIle 844
Db      4168 GCTAGCATATATCTATATTATGCTTCCATGAGCGCTGATTTGCGAAGCAATAGCA 4224
Qy      845 LeuArgArgGlyValLysAapValSerValThrGlnLeuGlnAlaLysAlaPheAspLeu 864
Db      4225 ACAGAAAAACCAATTAAAGAT-----AAGAAATTAGAGCTTAA-----TTA 4266
Qy      865 AlaAlaGlnValPheGlyArgTrpValAsp----- 874
Db      4267 GCTAAGGCTTATGAGTGCATTAAGAAAGATGACGATTAAGAAAAAGTAAATTCACA 4326
Qy      875 -----LeuLysGlnAapCysAapLysLysLeuGlnSerAapCysArgLysLysGlnAapCys 892
Db      4327 GCAGAACTTAAAGCAAGAAACAAATTAAGATAGACAAAGAT-----GTCGAGAAAGCTGCC 4380
Qy      893 LysAapLeuGlnGlnValCybLysLysIleAsnLys-----AlaCys 906
Db      4381 AAGAAATATCATGTAATCGCTTAAAGAACAAAAAGAAAGAGTGGGAAATTTGTAGAT 4440
Qy      907 ArgAsnLeuLysProLeuGlnValLysProHisGlnThrValThrGlnLysThrThr 926
Db      4441 GAAATGCTAATCCATGATGATGACAAAGAAAGAAAGCAAGAAAGCAAGTGAAGAAAGCCCT 4500
Qy      927 ThrThrThrThrThrThrThrValAlaAapProLysAlaThrGlnCybLysSerLeuGln 946
Db      4501 GTCAAGAGCGCTTATAGCGAAGAGATGCC----- 4533
Qy      947 ThrThrAapThrTrpValThrGlnThrSerThrHisThrSerThrIleThrSer 966
Db      4534 -----ACATTTGTTTAAAGCGCAATACACC-----CCCATTTGAATC 4569
Qy      967 ThrIleThrSerLysIleThrLeuThrSerThr 977
Db      4570 ACTCTGACTTCTTAAAGTAAATGCCACTCTCA 4602

```

RESULT 13

```

US-10-369-493-25273
: Sequence 25273, Application US/10369493
: Publication NO. US20030233675A1
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkle, Gregory J.

```

APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052) B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 25273
LENGTH: 5373
TYPE: DNA
ORGANISM: Saccharomyces cerevisiae
US-10-369-493-25273

Alignment Scores:
Pred. No.: 1,046-13 Length: 5373
Score: 274.00 Matches: 238
Percent Similarity: 35.04% Conservative: 193
Best Local Similarity: 19.35% Mismatches: 453
Query Match: 5.06% Indels: 346
DB: 18 Gaps: 51

US-10-654-416-14 (1-1023) x US-10-369-493-25273 (1-5373)

QY 6 LysArgGlnAlaValGlyThrGlnAsnSerIleAspGluGlnHisValIleuAlaLeuIle 25
DB 1996 AAAAAAGATTCTATTTCTCAAAAAGTTGATGATGAAGACAGATATTAATCCGAG 2055
QY 26 LeuIysGluAspGlyLeuSerGlu----- 33
DB 2056 CTGTGTAAGAACTGCTTTTCCAAAGTTTATTTACAGCACTTACTTACATTAATTTTAA 2115
QY 34 -----GlnGluCysLeuIysLeuIysLeuIysLeuIysLeuIysLeuIysLeuIysLeu 50
DB 2116 GAAACATATACAGATACGAACCTGATTTCTCACAGCCCTGACGAAAGAGCAATCAAT 2175
QY 51 LysLeuAsnIleGluGlnValHisArg-----LysLeuIysGlyPhe 64
DB 2176 AAAATATCTTTCGAGAAAGTCAAAAATTACAGAGCAATGACAAAATTGAAGGT--- 2232
QY 65 CysGluAspGlyLysAlaSerThrIysCysIysGluLeuIysAlaAsnIleGluIysLys 84
DB 2233 ---GAGATTAACCTTTTTCGAAACGAAACGAAACCCCATGAGAACCTCACCGAAAA 2289
QY 85 CysThrThrIleIysGlyLysLeuIysGlnAlaIleIysLysIleGlnIleIle--- 103
DB 2290 TTGATTGGCTTGACTTAATGACACAAAGG---TTGATGAGAAATACCAATTTTGAAAT 2346
QY 104 ---ThrAspLysAspCysLeuGluAsnGluGlnGlnCysLeuPheLeuGluGlyValCys 122
DB 2347 TCCTCACATTTCTGCTGTAAGAAAA---TTTCCATTTTG 2385
QY 123 SerLysGluLeuLys-----AspAspCysAsnThrLeuArgAsnLys 136
DB 2386 GAAACTGATTCGAAAGAACCTGACAGATTCTTGGATGAATGACCAACTGAGAGATGTA 2445
QY 137 CysTyrGlnLys---LysArgAspLysValAla----- 146
DB 2446 CTGGAACCTAAGGACAAAGAAATCAACTGCTTTTACCTGAGTACAAAGCAATCCAC 2505
QY 147 ---GlnGlnValIleuLeuArgLalaLeuArgSerAspLeuAsnGlySerValIleCysGlu 165
DB 2506 AAACAGAAAGACTCATATAAACTTTAGAAAAAGAGCTTGAAACATATTTTGTCTCAAAAG 2565
QY 166 LysLysLysLeuLysGluIleCysProValIleCysGlyArgGluSerAspGluLeuThrAsnLeu 185
DB 2566 AAAAAAGCAGAAAGATGCTATTAACAAATGGCTTAAGATTTATTCCTCTGAGTAGAGAG 2625
QY 186 CysLeuAsnGlnIleGluThrCysLysAsnIleLeuIleGluLysAsp----- 201

DB 2626 ATGCAGACGTTGAGAGAAATTGTAATAATTACAGAAAGAAAAAGATTAAGCAATGTC 2685
QY 202 -----LysLysCysGlyThrLeuLysThrAspValSerAlaValLeuGlySerPhe 218
DB 2686 AACCATCAGAAAGAGACTTAATCACTAAAGAAAGATTTGGCGCAAAATTAAGTAATA 2745
QY 219 Lys-----LysGluThrCys----- 223
DB 2746 AAAGCTATCAATGAAAACTGGAAAGAAATGAAATTCATATGTAATTAATTATCAAAAGAA 2805
QY 223 ----- 223
DB 2806 AAGGACATATTTGAGAGAACTGTTGAGTACAAATCCCGCTTTCAGAGTCATGACAAT 2865
QY 224 -----LeuGluLeuLeuGluGlnCysTyrPheTyrIleGlyAsnCysGlyAsp----- 239
DB 2866 CTAGTACGAAACCTAAGAAAAATTGAAATCTTACCAATTAATTAATTAAGATATGCA 2925
QY 240 -----AspAspIleIleIleCysAlaGlu----- 247
DB 2926 GCTGAAATAGTCTCTTAATTAAGCTGTAGAGATCAAAAAAGCAAGCATACAA 2985
QY 248 -----LeuGlyGlyLysCysGln----- 253
DB 2986 TTGCTTAATTTGCAAAATTAATTAATGATTCATGTCACAGAGAAAAAGAAATTTTCAATA 3045
QY 254 -----GlnGlnAsnIleAlaTyrMetProGly---ProAspPheAsp 267
DB 3046 GAAAGAGCAATATGAAAAATATTCACACAACTTAATAAAACATCTCGACTTGAATA 3105
QY 268 ProThrArgProGluAlaThrIleAlaGluAspIleGlyLeuGluGluPheTyrLys 287
DB 3106 CAAACGAAAGAGAAATTAATCTCAAAATCCGATTTCTCAAAAGATGAATAT--- 3156
QY 288 ValGluGluAspGlyValPheIleGlyLysAsnHisLeuArgAspAlaThrAlaLeuLeu 307
DB 3157 ---GAGTCCAGATTAGTCTCTTGAAAGAAAAATTAGAAAGGCGACAACA----- 3204
QY 308 AlaLeuLeuIleGluAspSerLeuLysLysAspAspLysGluLysCysGluGlu 327
DB 3205 ---GCAACGATGAAACGTGAATTAATCTCAGAAATTGACTTAAATACTAGGGA 3255
QY 328 AlaLeuGln----- 330
DB 3256 GAGCTGGAAGCCGAGTTAGCACTTATTAATAATCTCAAAATGAATTGAAACAAACTA 3315
QY 331 -----LysSerCysLysAsnProHisGluHis----- 339
DB 3316 GAACTTCAGAGAAAGCTTTGAAAGAGTTAAAGAAACGAGAGCAATTTAAAGAAAGAG 3375
QY 340 -----GluAlaLeuGluSerLeuCysLysLysValAsnGlyLeuSerAsn 353
DB 3376 AAAATTCAACTCGAAGAGAGCCACTGAAACCAACAGACATCAACAGCCTA---CGT 3432
QY 354 AspGlyThrLysLysCysGluGluLeuGlnAsnAspIleAsnLysThrCysLysIlePhe 373
DB 3433 GCCAATTTAGATCATTAAGAAAGAGCAGTAAATTTAGCAGCTCACTTGAAGAAAGTAC 3492
QY 374 ThrSerLysValThrAsn---AsnArgLeuPheAspPro-----ThrLysGlyAsnAsn 390
DB 3493 GAGAGCAAAATGGCAACAGAAAGGCAATATACAGAGAAATATCTCAATTAATATAT 3552
QY 391 GluIleValGlyTyrPglGlyLeuProThrPheLeuSerAsnGluAspCysAlaLysLeu 410
DB 3553 GAGATTACTTCC-----ACTCAACAGAAATTAATGAATCA----- 3585
QY 411 GluSerTyrCysPheTyrPheGluLysLysCysProAspGlyGluAsnAlaCysLysAsn 430
DB 3586 -----ATTAGAAAAAATAGACAACTGGAGGGCGAAAGTTAAACCA 3627
QY 431 IleArgAlaThr-----CysTyrLysArgGlyLysAspAla----- 442
DB 3628 ATGAAGACACTTCAGAGAGAAACAATTCAAATTTGAAAAAGTACAGAGATTGATTCCTTAAT 3687


```

Qy 443 -----ArgAlaenlyValleuGlnleuMetArgGlyMetLeuHieGly 458
Db 3688 TTGCMAATTAAAGCGTGAAGAAAGAGCAAGAAAGCAAGCAAGTATTGAGTCA 3747
Qy 459 SerAnlySerTLeuGluLysPheGlnGlnleuValLysGlu----- 476
Db 3748 -----ATCAAAAAGTGAATCAGAAACGGTAAAAATTAAAGATTACAA 3792
Qy 477 -----LysLeuLysLysGluLysGlySer 485
Db 3793 GACCAATGTAATTAAAGAAAGAAAGTGAATGGAATGGAACCAATCAGCGCTCA 3852
Qy 486 PheSerAnlySerTLeuPheLeuCyValGlnProAlaLysValAlaArgLeuLeu 505
Db 3853 GAAGATAGAACTTAAATATTG-----GAGCTACGAAGCAATCGAAGATGA 3903
Qy 506 ThrHisApleuAryMetLysThrLysPheLeuArgGlnleuApleuLysArgAsp 525
Db 3904 AAGAGAAAGTGAATGCTAAAAAGCACTGAATAAAATTCAATTGGAAGAAATGATCA 3963
Qy 526 Phe-----ProThrApleuLysAnlySerLysGluLeuGlyArgLysCyGlnApleuGly 543
Db 3964 TTATCTAAGGCCAAGAGAAATCAGAAATCTGATGTTGCTAGATTGAAGAAACATCTTCT 4023
Qy 544 GluApleuSerLysGluLeuThrTrpProCysHisThrLeuGlnGlnCysApleuLeu 563
Db 4024 GAGGAAAGAAA-----AATGCAAGAGGAGCAATGCGAAATTTGAAAAATTG 4062
Qy 564 GlyThrThrGluLeuLysGlnValleuLeuApleuHisLysApleuThrLeuLysAsp 583
Db 4063 AAAAATGAATATCAAAATCAAAATCAGCGCTCGAAAAAGAAAGAAAGCTTAAATGA 4122
Qy 584 GlnGluSerCysValLysTyrLysLysGluLysCysAnlySerTLeuArgArgGlyAsp 603
Db 4123 GGATCTTCACCAAT-----ACCAAGAAATATCTGAAAGATCAATACTTGGAAAGAT 4176
Qy 604 AsparGhrPheSerPheValCysValPheGlnApleuLysThrCysGluLeuMetValLysAsp 623
Db 4177 GAA-----TTAATTAGGCTACAAATAGAAAAC--GAATTTAAAGCCAAAGAG 4221
Qy 624 ValLysApleu--ArgCysGluValPheLysLysAnlyLeuValAserTyrLeuGlu 642
Db 4222 ATCATATAATACAGAGAGTGAAGCTAGAGAGTGAATTAACCAATGAT-----GAG 4272
Qy 643 PheLeuGluApleuAnlyThrAnlyLysLeuThrLeuGluArgAnlyCysProSerTLeuHis 662
Db 4273 CTTTAAAGAGAAACAAACACATATAAAAGCTTACAGATGAATTTCTTCATATAA 4332
Qy 663 ThrTyrCysApleuArgPheSerProAnlyCysProGluLeuThrLysGluAnlySerCysThr 682
Db 4333 GACAAAATTAATCAAAATAGCAAGAAAGCTTATCAATAGAGCAAGATTAAT----- 4383
Qy 683 LysLeuLysLysHisArgGluProPheTyrLysArgLysValleuGluApleuLys 702
Db 4384 -----AAACGAGACCTGAG--AGCTTGA 4407
Qy 703 ValGluLeuGlnGlyLysLeuThrApleuSerLysCysGluProAlaLeuLysArgTyr 722
Db 4408 GAACAACTCCGCGTGCACAAAGAAATCCAAAGCTAAAGTTGAAGAAAGGCTGAAAAA--- 4464
Qy 723 CysThrValAlaGlyApleuValApleuAnlySerLysSerGlyLeuCyLysAlaAnlyThr 742
Db 4465 -----CTT 4467
Qy 743 LysApleuAnlySerGlyLysApleuApleuLysArgLysGluLeuCyGluLysLeuVal 762
Db 4468 GAGGAAGAAATTTGAAA-----GAGGAAGCAAGAGCTTGAAGATCTAAAGAAATGATG 4521
Qy 763 LysGluValGluGluGlnCysLysValleuProThrGluLeuGlnProAlaLysAsp 782
Db 4522 AAAAAAGTAGAGTACGATTGAGAGCAATGAACGAGTTGAAATCTCTCAATGAAAC 4581

```

```

Qy 783 LeuLysLysApleuTyrLysThrTyrGluGluLysLysArgAlaGlu----- 799
Db 4582 ATTAAGAAATCAGATGAAATAATTAAGAACAGTCGAAAAAAGTGTGAGGATATTAAA 4641
Qy 800 ---AlaMetAnlySerSerLeuValleuSerLeuLysLysAnlyGluSerAnlyVal 818
Db 4642 AACTTGCAATGAGAAAGTGAATTTGATATCCGAAATCAACGAATCTGAAAGATATT 4701
Qy 819 SerLysSerAnlySerLys-----AnlyApleuLysAnlyAlaVal 831
Db 4702 GAAAGACTGAAAGTAAATCACTGAGGATGAAAGCAAAATCTGCTGAACTAGAAACCTGA 4761
Qy 832 SerAnlyLysLeuGlnApleuThrThrLysHisValLysLeuArgArgGlyValLysAsp 851
Db 4762 AAACAAAGCTTAATAACGCAACAAAGAAAATAAGATC-----AACGA 4806
Qy 852 ValSerValThrGluLeuGluAlaLysAlaPheApleuAlaGluValPheGlyArg 871
Db 4807 GAAGAAATACAGTTTGAATCTAAATTAAGATATAGAACGTGAACCTCAAGACAG 4866
Qy 872 TyrValApleuLysGluArgCysAnlyLysLeuGluSerApleu-----CysArgIle 888
Db 4867 CAAGCCGAATTAATAAGT-----AATCAAGAAAGAAAGAAATTAATTAATCTTCCGCTC 4920
Qy 889 LysGluApleuCysLysApleuGluGluValCysLysLysLeuAnlyLysAlaCysArgAsn 908
Db 4921 AAGGAATTAAGACGAAATTAATAGATAGACCAACAGAGGACAAAGAAATCCGAGAAAGAA 4980
Qy 909 LeuLysPro----- 911
Db 4981 CGTAGGCGCCAGTCAAGAAATTTCAAGTCAAAAGTCTCAATGATGAAAAAGCCATG 5040
Qy 912 ---LeuGluValLysProHisGluThrValThr-----Glu 922
Db 5041 TTGTTTAAATCAATTAACCAAGATTTGTAATTAAGACCAAGCAAGTGAAGAAAGGATGA 5100
Qy 923 SerThrThrThrThrThrThrThrThrThrThrValAlaApleuProLysAlaThrGluCys 942
Db 5101 GATACCTTTAAAAAGACACC-----GATTCAGAGCAAGAAATT 5142
Qy 943 LysSerLeu---GlnThrThrApleuThrTrpValThrGlnThrSerThrHisThrSerThr 961
Db 5143 GAAAGTTGCTAAAGAACTGATTAACCTGAAGCAAGCAAGCAATCTAATGTAAGAAAGACT 5202
Qy 962 SerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 981
Db 5203 AACCAAGATCGTTCCGAAATTTGATGACTGATGCTTTGTTACTGACTAGATGAGAAA 5262
Qy 982 ProThrLys-----CysThrThrGluGlu 990
Db 5263 AAGCCAAATATGCTCAAAAGCTGAAGATTTGGGCTTGAAATTAATTAATTCGATGAAAGAG 5322
Qy 991 ApleuApleuApleuValLysProSerGlu 1000
Db 5323 GATGATGAAGAAATGATGAAGAAAGAG 5352

RESULT 14
US-10-032-585-6646
: Sequence 6646, Application US/10032585
: Publication No. US20030180953A1
: GENERAL INFORMATION:
: APPLICANT: Terry, Roemer D.
: APPLICANT: Bo, Jiang
: APPLICANT: Charles, Boone
: APPLICANT: Howard, Bussey
: TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
: FILE REFERENCE: 10182-005-999
: CURRENT APPLICATION NUMBER: US/10/032,585
: NUMBER OF SEQ ID NOS: 8000
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 6646
: LENGTH: 5641

```

```

; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6646

Alignment Scores:
Pred. No.: 1,116-13 Length: 5641
Score: 274.00 Matches: 232
Percent Similarity: 37.28% Conservative: 168
Best Local Similarity: 21.62% Mismatches: 406
Query Match: 5.06% Indels: 267
DB: 17 Gaps: 48

US-10-654-416-14 (1-1023) x US-10-032-585-6646 (1-5641)

OY 26 LeuLYsgLUaSPgLYLeuSeRgLUngLUcybLYsLYLeuLYsLYeTyCySgLn 45
Db 2524 TTGAAAGAAAAAATTGATTCCACAGAACAGGCCAAAAAATATGAAATGCTATCAAT 2583
OY 46 GluLeuThr-----GluAlaLYLeuAaSnLle 54
Db 2584 AATATGACTAGAGACTTATTCATCTTAAAAAAGTGAAGCTGAAGCTCAAAATC 2643
OY 55 GluGlnValHtSaRgLYLeuLYsLYPheCySgLUaSPgLYLySaLaSPThrLYsCyS 74
Db 2644 AAACAAGCTGAAGAGAGATTAAAAATTGACTTATGAAATTTGAAAAACATAAAAAGCAT 2703
OY 75 LySgLUeLUySaLa--AaSnLleGULyLySgLYsThrLleLYSgLYLyLeuLYs 93
Db 2704 TATGATATTAACAATTAATTAATTTGAATTAATCAATTAATGAATTCAAACAGAAAAATAAT 2763
OY 94 GluAlaLeuLYsLYeLYsLleGlnLleLleThraP----- 105
Db 2764 GAA--TTGAGTAGAAAAATCGAAATCTTGACTGAAGATTAATTAATTAATGCCAAGCA 2820
OY 106 -----LYaSPCybLYsGLuSngLUngLUngLYs-----LeuPhe 117
Db 2821 CTGGAAGAGAACTTAAGATATACAGAAAGAAACATTTGATGATTAATTTGAGA 2880
OY 118 LeuGlnGLyValCYSeRSeLYsGLeU----- 126
Db 2881 TCTGCATCAGTTCATACATGACTTGAAGGCGAAATCAGATCTGAAGAGAAACA 2940
OY 127 -----LYaSPaSPCybAaSnThrLeuAaSnLYsCySgLYGlnLYsLYaRgSP 143
Db 2941 GTTAAAGCCAAAGAGAGTTGAAACACTTACACAGATTAAGATTGATTAATCTGAAAAAGAA 3000
OY 144 LySaValaGLngLUValLeuLeuAaRgaLaLeuAaGSeRSeSPLeU--AaSnLYSeRVal 162
Db 3001 TTGAAGAGAAACAACATCTAAAAAAATGAATTTGAAAGAGACATTCAAAAATTTACTGAT 3060
OY 163 lLeCySgLUyLYsLYeLUySgLUleCySProValMetGLyAaRgLUSeRSeSPGluLeu 182
Db 3061 TCTACCAATGAAGATTTAAGGAATTA-----GAGGATGAATTTG 3099
OY 183 ThrAaSnLeuCYbLeuAaSnGLuLYsGU-----ThrCYbLYsAaSnLleLUleGlnLYs 200
Db 3100 AAGACTATCAAAAAATCCAAATTAAGAAATCTCTCACAAAATTCGAATTAATTCAAAAA 3159
OY 201 AaPlyLYsCYbGLYThrLeuLYeThraSPValSeRValaLaLeuGLYSeRSePheLYs 220
Db 3160 TTGGAGAAA-----ACTGAAGAAAGATCTCCAGCA-----AAAGAT 3195
OY 221 GluThrCYbLeuGLuLeuLYsGLuGlnCYsLYrPheTYrLleGLYAsnCYsGLYAsp 240
Db 3196 GAAGAAATTTGAATTAATGAAGCTGAGACCAAAATCAAAATATGATTAATCTAAATTCAGAA 3255
OY 241 AaPLeuLeuLYsCYbLleGlnLeuGLYLySgLYsCYsGlnGLuAaSnLleAlaLYrMet 260
Db 3256 -----ATCTTAATTTTCAATCGAAGTTGAAGAAAGCAAGAAATCTCAT--- 3300
OY 261 ProPheGLYProAaSPRheAaSPProThraSPProGln--AlaThrLleAlaGLuSPLe 279
Db 3301 -----TCAAGTACCAAAAGATGAACACTTCTTCACTTTCTGAAGAAATCTC 3342

```

```

OY 280 ---GlyLeuGLuGLuPheTYrLYs----- 286
Db 3343 AAAAAATTGAAGAGAGATGAAGAAATACAAAACTTCAATGATAGCCAACTTTACAGA 3402
OY 287 LySaValGLuGLU-----AaPGLYValPheLleGLYLySaAaSnLleLeuAaRg 301
Db 3403 AAGATTAAGAGGACCAAGAAAGCCATGATGAATTTGAATTAAGACTTAACACATCCACC 3462
OY 302 AaPAlaThraLaLeuLeuAlaLeuLleLleGlnAaSPSeRSeLleUyLYsLYs----- 319
Db 3463 GAC-----TTACAGAGAAACATGCAAGCAAAAGTCACAA 3498
OY 320 -----AaPAPLYsGLYLySgLUngLUngLUngLUngLUngLYsSeRcyS 333
Db 3499 TTTGAGAGTGAACGAATGATATCAAGTCCAAATTTAGATGAAGCAATTAAGAAATTATCA 3558
OY 334 LySaAaSPProHtSaRgLUaLleGULeUgLUSeRSeLYsLYsLYsAaSnGLYLeuSeRSe 353
Db 3559 GATTAAT-----AGAGAAATAATTTCTTAATCTTGAAGAAAGAAACATGAATTCAT 3609
OY 354 AaPGLYThrLYsLYsCYbGLuGLuLeuGlnAaAaPLeaSnLYsThrCYbLYsLlePhe 373
Db 3610 AATTAACCTCAAAACACAGAAAGAAAGATTTCTGATTTGAAACTTCAGTGCATCTCT 3669
OY 374 ThreRLYsValThraAaAaRgLeuPheAaSPProThrLYsGLYAsnAaSnGLuLleVal 393
Db 3670 GAAGATTAAG--TCAAGAGCTTGAACACAGATTTGAAAGATTTGAAGAGAGAAATAATC 3726
OY 394 GLYrPrgLUgLYLeuProThrPheLeuSeRSeAaSnGLuAaSPCybAlaLYsLeuGLuSeRy 413
Db 3727 AAGTTAGAA-----ACAACTTAAGAGAAATAGAGAAACG----- 3762
OY 414 CybAaSPThrPheGLYLyS-----LYsCYsProAaSPGLY 424
Db 3763 -----ATGTTTGAAGAAAGAGACCAATTCGCAAGTTGAATGAACAAATGAAAGAGCTT 3816
OY 425 GluAaSnLaCYbLYsAaSnLleAaRgaLaThrCYsLYs-----Arg 438
Db 3817 GAACATGTTTGAAGAAATTAATCTGAACCAAGAAAGAAATTAATGATTAATTTGA 3876
OY 439 GlyLeuAaSP--AlaAaRgaLaAaSnLYsValLeuGlnLUaAaMeLAaRgLYMeLleUht 457
Db 3877 AAATTTAGAGCAGAAAGTCAATCATGATATCTAGCGCTTAAGAAATTTGCTCATTTAATT 3936
OY 458 GlySeRSeAaSnLYSeRTrLeuGLYLySgLUngLUngLUngLUngLUngLYsValCYsGLYLyS 477
Db 3937 GAAGACAGAAAGTGGAGTGAAGAAATGTCATCAAGTTGAATGAACAAATTTGAGAAA 3996
OY 478 LeuLYsLYsGLuAaSnLYsGLYSeRSeRSeAaAaPGLuLeuPheLleUeCYbValGln 497
Db 3997 CTCAAGGTGAAAGA-----GAGAAAGAAAGTCAAGGATATTTCA 4035
OY 498 ProAlaLYsAlaAaRgLeuLeuThraSPLeuAaSPLeuGMeLYsThrLlePhe----- 515
Db 4036 TCTGATTTGGCGCAAACTACCGATTGGGA--AAGATTAAACTACACTTGACAA 4092
OY 515 ----- 515
Db 4093 GTTTTGAAGAGAAAGATCTGAAGAAAGCTAAACAAGAAATGACATTTGAAG 4152
OY 516 -----LeuAaRgGLuGLuLeU-----AaPGLYLySaRgAaSPhe 526
Db 4153 AAAGAGTCAAAATCTGAAGAAAGAAATTTGTTGTTGGAAGATTCGAAAAAGACACAT 4212
OY 527 ProThr-----AaPlyAaSnCYbLYsGLuGLuGLYrLySgLYsCYbGLuAaPLeuGLY 544
Db 4213 ACCAACCAATTAAGAACTTCTGCTCAATTAAGAACTTAAGAACTTAAGACTCA 4272
OY 545 AaSPSeRSeLYsLleLleThrTrProCYbAaSPeThrLeuGLuGlnGLYCybAaSPRgLeu-- 563
Db 4273 ACCCAATGGAATTAAGAAAGACAGAGTGGAGTTGAAAAAAGTAAGAAATGAGCTAACT 4332

```


Db 553 TTAGCTAGAAAAATAAGAAAGAAA-----GAAAAGAAAAT 591
Qy 21 ValLeuAlaLeuIleLeuLys-----GluAspGlyLeuSerGluIn 34
Db 592 GTTTTAGTAGAGAAATTAAGGATATGAAAAATATATGATGTTCTTAAGAGAAAGA 651
Qy 35 -----GluCysLysLysLysLeuLys 42
Db 652 AGAGAGCTTTTGAAAAGAAAATATGATTTCTTTAATGAAAGCTTCAAAGAGATCTTAAG 711
Qy 43 TyrCysGlnGluLeuThrGluAlaLys-----LeuAsnIleGluIn 56
Db 712 GCAGAGAGGAAATACATAGAGAAAAGAGCTTTGGGCTTCCTCAATTTGAATATAGAGAA 771
Qy 57 ValHisArgLysLeuLysGlyPheCysGluAsp----- 67
Db 772 AAAAATAGATGCTCTTAAGATCTTATGAGAAAAGAAATGAAATAGACTTAAGAGAAA 831
Qy 68 -----GlyLysAlaAspThrLysCysLysGluLeuLysAlaAsnIleGlu 82
Db 832 AGAGCCCTTTTAGTGAAGATTCGTGAAAGCTTAAGCCTTACATTGATTAATTAAGAAAT 891
Qy 83 -----LysLysCysThrThrIleLysGlyLysLeu----- 92
Db 892 ACTTTAAACAAATTGATATATTAAGAACAAATCCTTCTCGTGAAGACACATGAAA 951
Qy 93 -----LysGluAlaIleLysLysLysIleGlnIleIleThrAspLysAsp 107
Db 952 GCTATTAAGCTTGAAAAAGAGCATGAGAAAAAATTAAGTAT----- 996
Qy 108 CysLysGluAsnGluInGlnCysLeu-----PheLeu 118
Db 997 GCAAAAGATTAATTAAGAAAAAGCACCTCTTAAGTTATGATTAAACATCATTAATATTA 1056
Qy 119 GluGlyValCysSerLysGluLeuLysAspAspCysAsnThrLeuArgAsnLysCysTyr 138
Db 1057 GATGCAATTAAGAAAGAAAGATCTTTTGATATATTAAG-----CTTGAAGAAAGAAAGCCTC 1113
Qy 139 GlnLysLysArgAspLysValAlaGluGluValLeuLeuArgAlaLeuArgSerAspLeu 158
Db 1114 CAAGGAAAAATTAAGAAAGCTTATTAAGAGCG-----AGT 1149
Qy 159 AsnGlySerValIleCysGluLysLysLysGluIleCysProValMetGlyArgGlu 178
Db 1150 AACAAAGAAATTAATTAAGAACAGAAACATTAAGATATA----- 1188
Qy 179 SerAspGluLeuThrAsnLeuCysLeuAsnGlnLysGluThrCysLysAsnIleLeuIle 198
Db 1189 ---GACTATTAACCTTGAATAATACAGAAATTTGGAGAGTAAAGATTGATTAATCTTAAGCTT 1245
Qy 199 GlnLysAspLysLysCysGlyThrLeuLysThrAspValSerAlaIleLysGlySerPhe 218
Db 1246 CCAGAAAGATTAAGAA----- 1260
Qy 219 LysLysGluThrCysLeuGlnLeuLysGlnCysTyrPheTyrIleGlyAsnCysGly 238
Db 1261 -----AATTAATTAATGAAGAAATTAATTTTAATGAAGATTAAT-- 1299
Qy 239 AspAspAspIleIleLysCysIleGluLeuGlyLysCysGlnGluGlnAsnIleAla 258
Db 1300 GATGAAAAATTAAGCAATTAAGAAATGAGCTTGACCTTGATGTAATTAATTTCAAGCT-- 1356
Qy 259 TyrMetProProGlyPheAspPheAspProThrArgProGluAlaThrIleAlaGluAsp 278
Db 1357 -----GATTTGAAAAAGCTTAAGTAAAAA----- 1383
Qy 279 IleGlyLeuGluGluPheTyrLysValGluGluAspGlyValPheIleGlyLysAsn 298
Db 1384 -----GAAATGCTTTTAATTAAGTTAGAGAAAG-----AGAAGT 1419
Qy 299 HisLeuArgAspAlaThrAlaLeuLeuAlaLeuLeuIleGlnAspSerSerLeuLysLys 318
Db 1420 AAATTAGATCACTACACAAAAAGCTTCAGAGATTTAAATTAAGAGCTTT-----CCT 1470

Qy 319 LysAspAsp-----LysGluLysCysGluGluAlaLeuGlnLysSerCys 333
Db 1471 AAAGATGATGACTCTTACATTTCAAGAAAAATTAATGATGTTCTAGCAAAAAATGGGCT 1530
Qy 334 LysAsnProHisGluHisGluAlaLeuGlnSerLeuCysLys-----LysAsnGlyLeu 351
Db 1531 AAGTATAGTGAGTATTAAGAAAGCTTAAGGCAAGCTTAAGGCTTGAAGAAATTCACAG 1590
Qy 352 SerAsnAspGlyThrLysLysCysGluGluLeuGlnAsnAspIleAsnLysThrCysLys 371
Db 1591 CAAGTTCTAAGAACCAAAAA---GAGGAAATGACAAATTAAGAGATTAATTAATCAAAA 1647
Qy 372 IlePheThrSerLysValThrAsnAsnArgLeuPheAspProThrLysGlyAsnAsnGlu 391
Db 1648 GTT-----AATATTAA 1659
Qy 392 IleValGlyTyrGluGluLeuProThrPheLeuSerAsnGluAspCysAlaLysLeuGlu 411
Db 1660 ATA-----GAGCTTTTAGAGAC-----GAA 1680
Qy 412 SerTyrCysPheTyrPheGluLysLysCysProAspGlyGluAsnAlaCysLysAsnIle 431
Db 1681 AACATGGCTCATGTAATTAAGAAAGCCTTAAGAGTGAAGG---GCATGT----- 1728
Qy 432 ArgAlaThrCysTyrLysArgGlyLeuAspAlaArgAlaAsnLysValLeuGlnGluAsn 451
Db 1729 ---CCTGATGT----- 1737
Qy 452 MetArgLysMetLeuHisGlySerAsnLysSerThrLeuGlnLysPheGlnGlnLeu 471
Db 1738 -----GGTTCTGTCATCATTAAGAAAGATTTAAGAAAGCTTGAC 1779
Qy 472 ValLysValCysGluLysLysLysLysGluAsnLysGly-----SerPhe 486
Db 1780 CTAAAGGCTTTGAAACACCTTAAGAGTGAAGCTCGAGCCTTGAAAAAAGCTAAGTTT 1839
Qy 487 SerAsnAspGluLeuPheIleLeuCysValGlnProAlaLysAlaIleArgLeuLeuThr 506
Db 1840 GAAATATGAAGAAATTA---GTTATGTGTGAAGCTTAAGCTTAAGTA----- 1881
Qy 507 HisAspLeuArgMetLysThrIlePheLeuArgGlnGlnLeuAspGlnLysArgAspPhe 526
Db 1881 ----- 1881
Qy 527 ProThrAspLysAsnCysLysGluLeuGlyArgLysCysGlnAspLeuGlyLysAspSer 546
Db 1882 ---GAGAAAAAATTAATTAAGAAAGCTTAATGAGACATTAATTAACCTTGAGAAAGACTTT 1938
Qy 547 LysGluIleThrTyrProCys-----HisThrLeuGlnGlnCysAsn 561
Db 1939 AAAGAGTTCTCTTGAAGATGAGAAAGAAATTTAATTAATTAATTAAGAAAGTAAC 1998
Qy 562 ArgLeuGlyThrThrGluIleLeuLysGlnValLeuLeuAspGluHisLysAspThrLeu 581
Db 1999 AAATTTAAT-----TTAGAAAAATTAACAATCTGATGATTAATATTAAGACCTTA 2046
Qy 582 LysAspGlnGlnSerCys-----ValLysTyrLeuLysGluLys----- 594
Db 2047 AGTGAAGCTCAATTAATTAATGAAGTTGAATATCAGAAAGAAAGACACTTGAAAGACAG 2106
Qy 595 CysAsnLysThrSerAlaArgGlyAsp-----AspArg 605
Db 2107 TGTGAAAA---CGTATTTGATTTAAAAAGCAATTGGAAGAGCTATTAAAGAA 2160
Qy 606 PheSerPheValCys---ValPheGlnAsnAlaThrCysGlnLeuMetValLysAspVal 624
Db 2161 TTATATGAAGATATACATTAAGAAATTTAAAGCTGAAGCTTAAGATTTCAAGATTTT 2220
Qy 625 Lys---AspArgCysGluValPheLysLys-----AsnIleLysAlaSerTyrIle 640
Db 2221 AAATTGAAATGAAGAAATTAATTTGGAAGAAAGAAAGATTAAGGCTTAAGAGCTGAA 2280

```
Qy 641 ILeGluPheLeuGluMetAsnThrAsnLeuIleThrThrLeuGluArgAsnCysProSer 660
    |||::: ||| |||
Db 2281 ATAAAGATCTTAAGAAACCTTTAAAC-----ATA 2310
Qy 661 TrpHisThrTyrcybaasnargpaseserproasnCyseProGlyLeuThrLyseGluAsnSer 680
    ||| ||| |||
Db 2311 AGACACACAGAAAAAGACAGCTAAATGATTAAGTGTAGTAACTAAAGAGAGAGCTTAGC 2370
Qy 681 CysThrLyseIleLys---LyseHisArgGluProPheTyrlsbaarglyseValaleuGluasp 699
    ||| ||| |||
Db 2371 AAAAATTAAGCTGAGCTTTAAAGAAAAAGATTAATAATTAACGAAATAATGAATATA 2430
Qy 700 AlaleuLyseValGluLeuGlnGlyLyseLeuThrasp---LyseSerLyseCyseGluProAla 718
    ||| ||| |||
Db 2431 AAAAATTAAGTAAAGAGATTAAGATTAATTTATACGAACTTAAGAAAAAGATTAAGCGTACA 2490
Qy 719 LeuLyseArgTyrcythrValAlaGlyAsnValAsnAsnAlaSerIleSerGlyLeuCys 738
    ||| ||| |||
Db 2491 ATTAATAATAATAAT----- 2502
Qy 739 LyseAlaAsnThrLyseAspAsnSerGlyLyseSeraspGluAspAlaArgLyseGluLeuCys 758
    ||| ||| |||
Db 2503 -----GAAAGAACGTAAATTATGT 2523
Qy 759 GluLyseLeuValLyseGluValGluGluGlnCyseLyseValaleuProThrGluLeuGln 778
    ||| ||| |||
Db 2524 GATTAAGAAATGAATGAATGAATGAACAAATACGAAGAGTGTAGTGAATAATTAAG 2583
Qy 779 ProAlaAlaAspLeuLys-----LyseAspTyrlsbaarglyseGluLeuLyseArg 796
    ||| ||| |||
Db 2584 TATCACAGTAAATTGTCCAGCTCAAGCATAGCAAAAGTT--AATGATATAGACAAATTA 2640
Qy 797 AlagluGluAlaMetAsnLyseSerSerLeuValleuSerLeuIleLyseLyseAsnGluSer 816
    ||| ||| |||
Db 2641 AATAAAATACTTATAGAAAGAAATTTGAAATATAGAGAAAGCTAAGAAATTAATCTA 2700
Qy 817 AsnValSerLyseSerAsnSerLyseAsnLyseAspLyseAsnAlaValSerAsnGlyLeuGln 836
    ||| ||| |||
Db 2701 AACGATTAAGAAATTAATTTGCTTAAGAAAGCATGTGAAAAAGTATAAAATGAATCTTCA 2760
Qy 837 AspThrThrLyseHisValLyseIleLeuArgArgGlyValLyseAspValSerValThrGlu 856
    ||| ||| |||
Db 2761 AAGGTAAATGAGCTGTGAGGCTTTTAAGCAAGACTTAAGAAATTAAGAAATTAACGAA 2820
Qy 857 LeuGluValaleuAspLeuAlaAlaGluValPheGlyArgTyrcValAspLeuLys 876
    ||| ||| |||
Db 2821 -----GAAAGCTGATGAAATTCAAAATAATATAGAGTTGAGAAAGCT 2862
Qy 877 GluArgCyseAsnLyseLeu---GluSerAspCyseArgIleLyseGluAspCyseLyseAspLeu 895
    ||| ||| |||
Db 2863 TCCAAAGCAAAAGCGCTTCAAGAAAGAAAGTATAAACTTGAGAGAGAGCTTAAACATA 2922
Qy 896 GluGluValCyseLyseLyseIleAsnLyseAlaCyseArgAsnLeuLyseProLeuGluValLys 915
    ||| ||| |||
Db 2923 GAAATAAAACTTAAGAACTTGGAACCTTCTTAACCTTAAGCAAGAGTTAGAACATATAA 2982
```

Search completed: October 29, 2005, 15:51:05
Job time : 2041 secs

This Page Blank (uspto)